

ЛЕГЕНДАРНАЯ КНИГА

МИРОВОЙ НАУЧНЫЙ БЕСТСЕЛЛЕР

Роджер ПЕНРОУЗ

О компьютерах, мышлении и законах физики

Новый УМ короля



Roger Penrose

THE EMPEROR'S NEW MIND

Concerning Computers, Minds and The Laws of Physics

Foreword by Martin Gardner

Р. Пенроуз

НОВЫЙ УМ КОРОЛЯ

О компьютерах, мышлении и законах физики

Перевод с английского
под общей редакцией
В. О. Малышенко

Издание третье



URSS
МОСКВА

ББК 22.12 22.3о 32.813 87.2

Пенроуз Роджер

Новый ум короля: О компьютерах, мышлении и законах физики. Пер. с англ. / Под общ. ред. В. О. Малышенко. Предисл. Г. Г. Малинецкого. Изд. 3-е. — М.: Издательство ЛКИ, 2008. — 400 с. (Синергетика: от прошлого к будущему.)

Монография известного физика и математика Роджера Пенроуза посвящена изучению проблемы искусственного интеллекта на основе всестороннего анализа достижений современных наук. Возможно ли моделирование разума? Чтобы найти ответ на этот вопрос, Пенроуз обсуждает широчайший круг явлений: алгоритмизацию математического мышления, машины Тьюринга, теорию сложности, теорему Геделя, парадоксы квантовой физики, энтропию, рождение Вселенной, черные дыры, строение мозга и многое другое.

Книга вызовет несомненный интерес как у специалистов гуманитарных и естественно-научных дисциплин, так и у широкого круга читателей.

«The Emperor's New Mind, with a new Preface from the Author»
was originally published in English in 1999.
This translation is published by arrangement with *Oxford University Press.*

Произведение «The Emperor's New Mind» с новым предисловием автора
впервые опубликовано на английском языке в 1999 г.
Перевод на русский язык публикуется по соглашению с *Oxford University Press.*

Перевод на русский язык осуществлен с английского издания 1999 г.

*Перевод и редакция: Андрей Дамбис, Юлий Данилов, Сергей Кокарев,
Виктория Малышенко, Игорь Ольшевский, Леонид Яковенко*

Издательство ЛКИ. 117312, г. Москва, пр-т Шестидесятилетия Октября, д. 9.
Формат 70×100/16. Печ. л. 25. Зак. № 3042.

Отпечатано в ООО ПФ «Полиграфист». 160001, г. Вологда, ул.Челюскинцев, д.3.

ISBN 978-5-382-00744-1





О серии

Издательство УРСС продолжает серию книг «Синергетика: от прошлого к будущему».

Синергетика, или теория самоорганизации, сегодня представляется одним из наиболее популярных и перспективных междисциплинарных подходов. Термин синергетика в переводе с греческого означает «совместное действие». Введя его, Герман Хакен вкладывал в него два смысла. Первый — теория возникновения новых свойств у целого, состоящего из взаимодействующих объектов. Второй — подход, требующий для своей разработки сотрудничества специалистов из разных областей.

Но это привело и к замечательному обратному эффекту — синергетика начала оказывать все большее влияние на разные сферы деятельности и вызывать все больший интерес. Сейчас этим подходом интересуются очень многие — от студентов до политиков, от менеджеров до активно работающих исследователей.

Синергетика прошла большой путь. Тридцать лет назад на нее смотрели как на забаву физиков-теоретиков, увидевших сходство в описании многих нелинейных явлений. Двадцать лет назад, благодаря ее концепциям, методам, представлениям были экспериментально обнаружены многие замечательные явления в физике, химии, биологии, гидродинамике. Сейчас этот междисциплинарный подход все шире используется в стратегическом планировании, при анализе исторических альтернатив, в поиске путей решения глобальных проблем, вставших перед человечеством.

Название серии «Синергетика: от прошлого к будущему» тоже содержательно. Как говорил один из создателей квантовой механики, при рождении каждая область обычно богаче идеями, чем в период зрелости. Видимо, не является исключением и синергетика. Поэтому мы предлагаем переиздать часть «синергетической классики», сделав акцент на тех возможностях и подходах, которые пока используются не в полной мере. При этом мы надеемся познакомить читателя и с рядом интересных работ, ранее не издававшихся на русском языке.

«Настоящее» — как важнейший элемент серии — тоже понятно. В эпоху информационного шума и перманентного написания то заявок на гранты, то отчетов по ним, даже классики синергетики очень немного знают о последних работах коллег и новых приложениях. Мы постараемся восполнить этот пробел, представив в серии исследования, которые проводятся в ведущих научных центрах страны.

«Будущее...» — это самое важное. От того, насколько ясно мы его представляем, зависят наши сегодняшние усилия и научная стратегия. Прогнозы — дело неблагодарное, — хотя и совершенно необходимое. Поэтому ряд книг серии мы надеемся посвятить и им.

В редакционную коллегию нашей серии любезно согласились войти многие ведущие специалисты в области синергетики и нелинейной динамики. В них не следует видеть «свадебных генералов». В их задачу входит анализ развития нелинейной динамики в целом и ее отдельных областей, определение приоритетов нашей серии и подготовка предложений по изданию конкретных работ. Поэтому мы указываем в книгах серии не только организации, в которых работают эти исследователи, но и важнейшие области их научных интересов.

И, конечно, мы надеемся на диалог с читателями. При создании междисциплинарных подходов он особенно важен. Итак, вперед — в будущее.

Председатель редколлегии
Г. Г. Малинецкий

От редколлегии серии.
Синергетика, нелинейность и концепция Роджера Пенроуза

В этих заметках мне бы хотелось высказать и аргументировать две мысли, касающиеся замечательной книги Роджера Пенроуза «Новый ум короля».

Первая состоит в том, что мы имеем дело с удивительной книгой. Возможно, с книгой, открывающей новый жанр научной литературы. И если будущее науки состоится, то, вероятно, таких книг будет становиться больше и больше.

Вторая мысль связана с синергетикой. Хотя это слово не упоминается в книге, но подход, стиль, постановка проблем и предлагаемые пути их решения идут от синергетики и, более широко, от нелинейной динамики.

Несмотря на то, что эта книга написана в 1989 году — 15 лет назад — и переведена на русский язык только в 2002 году, на мой взгляд, в ней намечены контуры будущего синергетики. Мы с вами знаем, что будущее, как правило, неединственно. Поэтому правильнее будет сказать, что в книге предложен вариант развития нелинейной науки и междисциплинарных исследований в области фундаментальных наук. По-моему, вариант очень интересный, и к нему стоит привлечь внимание нашего синергетического сообщества. Стоит сделать это именно сейчас, когда интерес к тому подходу, который мы развиваем, применяем или которым просто интересуемся, стремительно растет.

В начале лета наша «Синергетика: от прошлого к будущему» вышла на новый рубеж. В России сейчас ежемесячно продается более 1000 книг, изданных в нашей серии. Это важная веха. Много серьезных, интересных работ по синергетике издается и другими издательствами, и издательством УРСС вне серии. Популярность синергетики растет. Поэтому прежде чем количественные изменения перейдут в качественные, прежде чем облик науки неузнаваемо изменится, стоит подумать о перспективах. И книга Роджера Пенроуза дает богатую пищу для таких размышлений.

Новый жанр научной литературы

Но роза пахнет розой,
Хоть розой назови ее, хоть нет.

В. Шекспир

Мартин Гарднер — признанный классик научно-популярного жанра, — написавший предисловие к этой книге, трактует эту работу как «популяризаторскую», как «понятную не только профессионалам». Да и сам Роджер Пенроуз рассматривает ее как «свою первую попытку написать научно-популярное произведение».

С этим взглядом трудно согласиться. Посудите сами. Взглянем «с птичьего полета» на жанр научной популяризации. И сразу станет видно несколько больших групп книг.

Книги, посвященные выдающимся научным достижениям. Помнится, мне в детстве попадалась книга М. Гарднера «Теория относительности для миллионов». И ведь действительно для миллионов. Еще двадцать с небольшим лет назад, в добрые старые времена тираж журнала «Знание — сила» был около миллиона, а «Наука и жизнь» издавалась трехмиллионным тиражом... Книги и журналы этого жанра максимально наглядно и доступно рассказывают о том, что полезно и интересно знать многим. В упомянутой книге обсуждалась относительность пространства, времени, одновременности, удивительный парадокс близнецов, знаменитая формула $E = mc^2$, имеющая непосредственное отношение к ядерному оружию. Множество книг такого сорта было выпущено «Молодой гвардией» в серии «Эврика».

Книги, посвященные основам наук. «Занимательная физика», «Занимательная математика», «Занимательная астрономия» и прочие, прочие, прочие. Королем этого жанра был Л. И. Перельман. Главный адресат таких книг — школьники.

Именно им надо показать, что наука — это интересно, увлекательно и доступно.

Книги развлекательного типа. Это разнообразные «математические досуги», «лаборатории дома» и т. д. Множество блестящих книг такого жанра перевел «поэт нелинейной динамики» Юлий Александрович Данилов. К примеру, тот же Мартин Гарднер, «Математические головоломки и развлечения» [1]. И когда есть что-то понятное, но и вместе с тем неожиданное, а также талантливый и увлеченный всем этим автор, то читатели всегда найдутся. Как правило, это книги без формул и без ссылок, рассчитанных на дотошного читателя, готового дойти до первоисточников.

Научно-художественная литература. Эти книги не только об идеях, теориях и результатах, но и о творцах науки. Например, такова книга выдающегося популяризатора науки Даниила Семеновича Данина «Неизбежность странного мира» [2].

Так вот, книга Роджера Пенроуза *не относится* ни к одному их этих жанров. Судите сами. Нового крупного достижения, которому посвящена была бы книга, *нет*. Напротив, есть несколько тупиков, из которых исследователям, занимающимся фундаментальными задачами, уже не первое десятилетие не удается выбраться. Так что книга посвящена не достижениям, успехам и перспективам, а неудачам и поискам выхода из нынешней проблемной ситуации.

Хотя в книге объясняется, что такое вектор, комплексные числа, фазовое пространство, метрика и многое другое, это книга *не об основах наук*, а об их вершинах.

Хотя формул действительно немного и есть очаровательное обращение к читателям «Как читать математические формулы» (после которого были бы уместны любые выкладки), это — *не развлекательная книга.*

Наконец, это и не научно-художественная работа. Ведь в ней идет речь не о вдохновении творцов, а о красоте идей, составляющих фундамент современного естествознания.

Так что же это? И кто читатели этого произведения? Кому адресована эта книга?

Представьте себя в положении ученого, намеревающегося организовать междисциплинарные исследования глубокой и интересной проблемы. Здесь далеко до результатов. Здесь нет еще исследовательской программы, под которую можно было бы организовать институт или получить грант. Здесь нужен поиск, в котором бы участвовали представители многих научных дисциплин, совместные усилия. Определяя синергетику, Герман Хакен в свое время писал именно об этом — о междисциплинарном подходе, разработка которого требует усилий представителей многих научных дисциплин.

Что делать в этом случае? Обратиться к коллегам по цеху? Скорее всего, они вас вежливо выслушают, пожмут плечами и вернутся к своим прежним научным делам. Междисциплинарный диалог — дело нелегкое, и трудно надеяться, что многие к нему окажутся готовы. Обратиться к профессионалам из «чужих» областей? Тут шансы на успех еще меньше. Кроме того, диалог затрудняет в лучшем случае «языковый барьер». А в худшем — предрассудки иного научного цеха оказываются просто непреодолимы. Выращивать студентов и аспирантов? Но это тоже работа на многие годы, а начать хотелось бы уже сегодня.

Как же действовать? После появления книги Роджера Пенроуза это ясно. Надо написать книгу, рассчитанную на будущих соратников по предполагаемым исследованиям. Р. Пенроузу удалось написать удивительную книгу. Она не имеет, на мой взгляд, аналогов в научной литературе и, очевидно, не является научно-популярной.

Во-первых, потому что это скорее интуитивная, чем рациональная книга. Она посвящена догадкам, гипотезам, аналогиям, а не полученным результатам. И в начале исследований она и не может быть другой! И, конечно, создание такого текста требует и незаурядного мастерства, и большой интеллектуальной смелости.

Во-вторых, в этой книге своим будущим коллегам надо максимально ясно и просто изложить результаты той дисциплины, в которой сам автор является ведущим специалистом и которые могут оказаться важными для них. (Это и делает книгу внешне не похожей на научно-популярную.)

В-третьих, придется наметить эскиз всей междисциплинарной программы и сказать, каким видится решение поставленной проблемы.

В-четвертых, нужно рассказать о множестве разных результатов и идей в тех областях, в которых автор не является узким специалистом. Поэтому понадобится большая библиография, чтобы последователи смогли уточнить детали и двинуться дальше, а также ясная собственная оценка альтернативных точек зрения (в научном мире, да, наверно, и в других мирах, это не просто).

В-пятых, это надо сделать кратко и гармонично, чтобы хотя бы некоторые читатели смогли добраться от начала до конца.

Роджеру Пенроузу все это удалось. И если междисциплинарные исследования будут успешно развиваться, то книг в этом замечательном жанре будет появляться все больше.

Сверхзадачи науки и научная картина мира

Я хотел, чтобы большинство из вас смогло оценить красоту нашего прекрасного мира и вместе с тем получить физическое представление о мире, которое, я думаю, составляет сейчас главную часть культуры нашей эпохи.

Р. Фейнман

Мне и моим коллегам не раз доводилось говорить и писать [3], что будущее науки в большой мере зависит от того, насколько нынешнее и будущее поколение исследователей преуспеет в решении трех задач.

Разработка теории управления риском. Важнейшей задачей ученых стал прогноз бедствий, катастроф, нестабильностей, кризисов в природной, техногенной, социальной сферах. И, естественно, разработка и анализ стратегий и технологий, которые позволяли бы парировать нынешние и будущие угрозы. Наверно, это и есть научная основа концепции устойчивого развития, о которой в последние годы много говорят.

Нейронаука. Человек является главной загадкой, которую наука XX века оставила в наследство исследователям нашего столетия. Если прошлый век был веком высоких технологий в промышленности, сельском хозяйстве, военном деле, сфере управления, то XXI веку, вероятно, суждено стать веком высоких гуманитарных технологий. Технологий развития и использования различных индивидуальных и коллективных возможностей человека. Важнейшим объектом исследования в этом направлении, конечно, будет мозг. В этом контексте книга Р. Пенроуза очень показательна. В самом деле, в ней выдающийся специалист по математической физике утверждает, что понимание принципов работы мозга требует использования огромного научного арсенала от теории вычислений и электродинамики до общей теории относительности. И более того, что в арсенале теоретической физики пока есть *физические* пробелы, не позволяющие осмыслить феномен сознания.

Теоретическая или альтернативная история. Направление, позволяющее, опираясь на методы естественных и гуманитарных наук, на компьютерное моделирование анализировать исторические альтернативы. Это направление может превратить стратегическое планирование из искусства в науку. Помочь узнать, какова цена того или иного выбора, который делают цивилизации или человечество в целом.

Но, как стало особенно ясно в эпоху нынешнего кризиса науки, общество ждет от исследователей не только этого. Оно остро нуждается в *научной картине мира.* Той самой картине, которая должна быть важной частью общей культуры нашей цивилизации.

Отсутствие такой картины ставит под угрозу смыслы и ценности. Оно будет означать откат к религии, мифологии, суевериям... Оно может резко уменьшить те шансы, которые остались у человечества.

Создание целостной картины мира — осознание единства научного знания и очерчивание тех проблем, которые нужно решить — важная междисциплинарная задача. Это именно та задача, с которой, вероятно, в ближайшие десятилетия придется иметь дело синергетике. Директор Института философии РАН академик В. С. Степин считает, что синергетика будет в центре естественно-научной парадигмы XXI века [4]. Так что, видно, пришла пора людям, занимающимся синергетикой, всерьез поразмышлять о своей нынешней и будущей ответственности. Выскажу крамольную мысль: научная картина мира — слишком серьезная и ответственная вещь, чтобы сейчас ее создание и осмысление можно было бы отдать на откуп только философам. И, по-моему, главный пафос книги Роджера Пенроуза это не парадигма искусственного интеллекта, не белые пятна квантовой механики, не тайны сознания. Это попытка очертить контуры естественнонаучной картины мира.

Назовем вещи своими именами. Исследования в области физики элементарных частиц, космологии, теории эволюции, а также во многих-многих других областях весьма вероятно не приведут с созданию новых видов оружия, к новым товарам и услугам. И все эти работы, как правило, довольно дорогие, видимо, никогда себя не окупят. Но если мы хотим узнать пределы своих возможностей познавать, хотим работать вблизи этих пределов и не хотим утратить способности к познанию, то такие работы надо вести. Вопрос в приоритетах, сроках, ресурсах, в структуре нашего незнания. Здесь опыт, уже накопленный синергетикой, и ее представления были бы очень важны.

И тут философии науки (если таковая сохранится в будущем) придется быть гораздо ближе к науке. Может быть, это возврат к Платону и Аристотелю, считавшим, что философам надо детально представлять геометрию и другие дисциплины — научную конкретику того времени? В противном случае неизбежна вторичность, отсутствие конструктива, который ученые ищут в философии, после того как кандидатский минимум, институтские и школьные экзамены по этому предмету сданы.

И с этой точки зрения, с позиций философии науки (что бы это не значило), книга Роджера Пенроуза является выдающейся.

С полвека назад один из создателей квантовой механики Эуген Вигнер написал замечательную статью, посвященную пределам науки [5]. В ней он поставил под сомнение саму будущность научного познания. (Думаю, что само развитие синергетики можно рассматривать как своеобразный ответ на несколько поставленных в этой статье вопросов.)

Обсуждая научную картину, в создании которой исследователи XX века, по мнению Э. Вигнера, не преуспели, он приводит очень интересную и содержательную метафору. Он сравнивает здание науки с замком. Целостная научная картина, с этой точки зрения, означает, что некто, обладающий временем и силами, мог бы попасть из любой точки этого величественного здания в любую другую, не выходя из него. Другими словами, разные разделы науки в идеале должны были бы иметь взаимосвязи, позволяющие переходить от одних научных дисциплин, концепций, категорий к другим, не опираясь на домыслы, мифы или благие пожелания.

По мнению Э. Вигнера, современная наука не такова. В ней есть две дисциплины, претендующие на осмысление мира во всей полноте. Это физика и психология.

Физика — потому что именно она начинает с микромира и стремится выводить законы следующего уровня организации материи, исходя из законов предыдущего. В самом деле, из законов квантовой механики следует уравнение Шредингера, а с ним и набор всех химических элементов с их свойствами. Квантовая химия — способы анализа этого уравнения в более сложных случаях — уже сейчас позволяет многое сказать о динамике достаточно сложных химических реакций. Ну а живые существа — объект биологии — это прежде всего химические машины. И так далее — по этой лестнице вверх. Не всегда удается далеко продвинуться, но пафос именно таков, да и успехи впечатляющие. Роджер Пенроуз делит все теории на превосходные, полезные, пробные и тупиковые: «Для того, чтобы теорию можно было причислить к разряду ПРЕВОСХОДНЫХ, совершенно не обязательно, по-моему, требовать от нее полного согласия со всеми явлениями в мире — однако диапазон явлений и точность их описания должны быть... феноменальными». И сегодня физика является главным источником понимаемых таким образом превосходных теорий.

Все, что мы знаем об окружающем мире, в конечном итоге — результат наших ощущений и работы сознания. И цель наших усилий, в конце концов, состоит в том, чтобы сменить одни ощущения на другие. Поэтому психология, понимаемая как наука о сознании, об ощущениях, о нашей «виртуальной реальности», должна была бы быть центральной.

И если между двумя главными башнями замка — физикой и психологией — нет моста (психология не используется для понимания физических феноменов, а физика довольно слабо применяется для решения частных психологических задач), то нет и картины мира.

С легкой руки физика и писателя Чарльза Сноу, стали популярны сетования на растущую пропасть между естественно-научной и гуманитарной культурами. Попытка преодолеть эту пропасть стала одним лейтмотивом развития синергетики. Попытки преодолеть, опираясь на модели, применимые и в том, и в другом мире.

И радикальность концепции Роджера Пенроуза в этом контексте вызывает глубокое уважение. И я бы даже сказал, ожидание чуда. Таковые иногда бывают в науке.

В самом деле, на основе глубокого анализа теории вычислений, собственного математического творчества, «превосходных» физических теорий и нейробиологических экспериментов выдвинута смелая, и наверно правильнее сказать, революционная гипотеза.

Гипотеза о том, что в старой доброй квантовой теории, которую студенты проходят уже более 70 лет, нет принципиального элемента. Элемента, который позволяет объяснить феномен сознания. Из учебников квантовой механики известно, что именно наблюдатель «превращает» волну, ψ-функцию, комплексную амплитуду в частицу. При этом говорят о редукции волнового пакета, которую Р. Пенроуз для краткости называет R-процедурой. R-процедура — «больной ребенок» квантовой механики. Это одна из тем дискуссии Бора с Эйнштейном. Это источник «принципа дополнительности». (Может быть, у некоторых физиков и философов есть теплое отношение к этому принципу. Но мне он кажется попыткой возвести непонимание и пробелы теории в ранг добродетели.) Наконец, это один из немногих процессов, по поводу которого физики «приговаривают» и еще не умеют писать интересные содержательные формулы. Это неисчерпаемый источник для многочисленных интерпретаций квантовой механики.

Но, по мнению Роджера Пенроуза, феномен сознания тесно связан с физическими процессами, происходящими на квантовом уровне, и, в частности, с этой самой R-процедурой. Но не той «субъективной», которую наблюдают и знают физики, а с новой — так называемой *объективной редукцией*. Если квантовую систему не наблюдают, то она, по мнению Р. Пенроуза, тем не менее, может претерпеть «редукцию» — проявить макроскопические

свойства и из множества возможностей выбрать одну. Это новое физическое явление, предсказанное Р. Пенроузом, будучи обнаружено, имело бы огромное значение.

В частности, мировоззренческое значение. Оно могло бы дать тот самый мост, который соединил бы две твердыни научного замка, два взгляда на мир — физический и психологический.

Программа или проект Роджера Пенроуза, связанный с поиском физических основ сознания, на этом этапе оказалась большой удачей. Не только его самого, его идеи, но и самого междисциплинарного подхода. Благодаря этой книге появились биологические исследования («нанобиология») и кандидаты в те структуры в нервных клетках — нейронах — в которых могла бы происходить предсказанная Р. Пенроузом «объективная редукция». Появились сильные оппоненты, которые позволили отточить одни аргументы и отказаться от других. Это видно по следующей книге «Тени разума» [6], которую недавно также перевели на русский язык.

Наконец, эта книга является поводом для оптимизма и для всех отечественных читателей, не равнодушных к судьбам российской науки и образования. В самом деле, тиражи научно популярных книг упали за последние 15 лет в сотни, а в ряде случаев в тысячи раз. Поэтому был очень приятен огромный интерес читателей к этой книге и к книге Б. Грина «Элегантная Вселенная» [7]. Отчасти этот успех является неожиданным. Ведь обе книги относятся к естественно-научному мировоззрению и очень далеки и от сиюминутных забот и от практических задач, в решении которых могла бы помочь наука.

Значит, у нас есть много людей, готовых смотреть на звезды и размышлять про загадки Вселенной. И это прекрасно.

Новый наряд короля

Вы полагаете, все это будет носиться?
Я полагаю, что все это следует шить!

Ю. Левитанский

Иногда выдающиеся теории вырастают из попытки решить частную задачу или ответить на конкретный вопрос. Например, целью фундаментального труда Евклида было разработать строгую теорию правильных многогранников (помните из школы — куб, тетраэдр, октаэдр, икосаэдр и додекаэдр?).

Так же действует и Р. Пенроуз, ставя себе целью показать, что наше сознание *принципиально* отличается от всего того, что сейчас может быть смоделировано на компьютере. Когда в книге говорится про «невычислимость», то имеется в виду именно это.

Когда речь идет о новой теории, тем более междисциплинарной, то часто возникает серьезная трудность. Именно та, про которую писал Г. Х. Андерсен в замечательной сказке «Новый наряд короля». Тогда, помнится, портные просили экзотических тканей, дорогих нитей и много другого. Но придворные, да и сам король, не представляли, какие же материалы на самом деле понадобятся.

На мой взгляд, здесь есть прямая аналогия с научным творчеством. Новое приходится «шить» из старого, которое, скорее всего, не очень «подходит», не очень соответствуют замыслу. Да и риск услышать «А король-то голый!» даже при самом добросовестном отношении к делу тоже достаточно велик.

Итак, из чего же «сшита» концепция Р. Пенроуза? Из каких результатов, идей и гипотез? Посмотрим на материал, не отделяя «новое» от «старого» и «простое» от «сложного».

Сильная трактовка искусственного интеллекта утверждает, что мозг подобен своеобразному компьютеру. И все, что сейчас доступно мозгу, в частности сознание, также со временем станет доступным компьютерам. Сознание, как учит философия и психология, — многогранный феномен. Р. Пенроуз сосредоточивает внимание только на одной из сторон — на творчестве, и даже более узко — на математическом творчестве. Итак, по мнению Р. Пенроуза, компьютеры никогда не обучатся «пониманию», у них никогда не будет «озарения» и «творчества».

Могущество математики в формализации, в осмыслении алгоритмов, в определении собственных пределов наглядно иллюстрируют результаты Тьюринга и Геделя. Результат Тьюринга (универсальная машина Тьюринга) сводится к теоретическому доказательству возможности построения компьютера. То есть некоторой машины, которая способна по данному входному числу (исходные данные) и характеристике алгоритма («программе») получить результат. При этом машина оказывается не «узко специализированной» — сделанной для выполнения данного алгоритма, а универсальной — способной исполнять любые алгоритмы.

С другой стороны, конструкция Тьюринга позволяет решить проблему Гильберта об алгоритмической разрешимости. То есть показать, что *не существует* некой механической процедуры для решения всех математических задач, принадлежащих к некоторому широкому, но вполне определенному классу.

В давние времена Ренэ Декарт писал трактат об универсальном методе решения математических задач. Даже если считать, что такой «метод» — решение задач на компьютере с неограниченной памятью и быстродействием — существует, то и для такого метода есть неразрешимые задачи.

Более того, результаты Геделя показывают, что дело не в пороке «компьютерного подхода», а в самой природе математики. Гедель показал, что *любая* достаточно широкая формальная система аксиом и правил вывода, свободная от противоречий, должна содержать утверждения, которые нельзя ни доказать, ни опровергнуть в рамках этой формальной системы. Итак — не формализм, а интуиция во многом определяет развитие математики. «Глубина», «оригинальность», «красота». «соответствие реалиям нашего мира» — такие интуитивные или «экспериментальные» критерии приходится относить к самым строгим математическим результатам.

Казалось бы, это именно то, что теориям не удается осуществить с помощью алгоритма, что отличает искусственный интеллект от естественного. Но Р. Пенроуз придерживается еще более радикального, а может быть, более правильно сказать, романтического взгляда: «„Богом данные" математические идеи существуют в определенном смысле вне времени и независимо от нас смертных... Настоящая математическая истина выходит за пределы сотворенного человеком». По мысли автора, «мир идей» Платона это та реальность, с которой исследователи имеют дело в минуты творчества.

В качестве «неалгоритмичности» нашего сознания Р. Пенроуз рассматривает несколько замечательных примеров из «нерекурсивной математики». Эта часть математики имеет дело с задачами, ответ на которые может быть «да» или «нет», однако не существует общего алгоритма, позволяющего выяснить, какой ответ верен. Другими словами, для этих задач нет формализма, «станка», позволяющего сэкономить усилия. Каждая задача требует «ручной работы».

Блестящий пример таких проблем дают задачи с замощением плоскости без наложений и перекрытий с помощью заданного набора плиток. В частности, «геометрической сенсацией» XX века стало обнаружение набора плиток, которыми можно замостить плоскость *только непериодическим образом.* Прогресс в этом направлении оказался грандиозен. Если в 70-х годах Бернер смог предъявить набор 20 426 плиток различной формы, то для решения этой задачи Р. Пенроуз в конце концов сумел сократить число типов плиток до двух! Конечно, этот результат вошел во все учебники современной геометрии.

Эта история имеет счастливое продолжение. Вдохновленные непериодическими замощениями плоскости физики-экспериментаторы обнаружили *квазикристаллы* — объекты, занимающие промежуточное положение между упорядоченными и неупорядоченными твердыми телами.

Большое впечатление производит и обсуждение «превосходных» в смысле Пенроуза физических теорий. Рассматривая квантовую механику, автор формулирует свое «предчувствие» будущей теории в следующих словах: «Я действительно считаю, что вопрос о свободе воли представляет собой реальную, а не вымышленную проблему — но она в высшей степени нетривиальна и ее трудно сформулировать адекватно. Вопрос о *детерминизме* в физическом смысле важен, однако я убежден, что он не является камнем преткновения. Например, мир может быть детерминистическим, но *невычислимым.* Иначе говоря, будущее может определяться прошлым, но только рассчитать его при этом будет в *принципе* невозможно... Соответственно свобода воли, которой мы наделены (по нашему глубокому убеждению), должна быть тесно связана с какой-то невычислимой составляющей законов, управляющих тем миром, в котором мы живем».

Из квантовой механики автор выбрал классические проблемы — шредингеровскую кошку и популярный ныне парадокс Эйнштейна—Подольского—Розена. В первом случае речь идет о странностях, возникающих при усилении квантовых свойств («перепутанные состояния», суперпозиция, линейность) до макроскопического уровня. В конечном счете о проблеме редукции волнового пакета. Во втором — о том, что из первых принципов квантовой механики следует, что измерение, проведенное над одной частицей, должно немедля определить состояние другой. И ни соответствующего поля, ни механизма, ни разумных формул, помогающих понять это, пока не видно. И название главы «Квантовая магия и квантовое таинство» весьма красноречиво. По мысли Р. Пенроуза, это именно тот уровень масштабов и те процессы, анализируя которые, учитывая их влияние на другие уровни организации, следует искать корни «невычислимости» нашего мира в целом и сознания в частности.

Очень любопытным представляется взгляд Р. Пенроуза на необратимость, на стрелу времен, на второе начало термодинамики.

С одной стороны, именно второе начало запрещает существование вечных двигателей второго рода и проходится в курсе элементарной физики. Кажется, Эрвин Шредингер сравнил второе начало термодинамики с всесильным директором, который определяет направление всех процессов, а закон сохранения энергии со скромным

бухгалтером, скрупулезно учитывающим приход и расход. А с другой...

Как ни странно, со времен Клазиуса, который ввел понятие энтропии, это понятие вызывает и интерес, и дискуссии. В добрые старые времена, когда я готовился к экзамену по статистической физике, меня поразила формулировка из знаменитого курса теоретической физики Л. Д. Ландау и Е. М. Лифшица. В этом курсе, отличающемся краткостью, ясностью и точностью, следующие строки выглядели загадочно: «...квантовая механика в действительности существенным образом содержит неэквивалентность обоих направлений времени. Эта неэквивалентность проявляется в связи с основным для квантовой механики процессом взаимодействия квантово-механического объекта с системой, подчиняющейся с достаточной степенью точности классической механике...

Таким образом, в квантовой механике имеется физическая неэквивалентность обоих направлений времени, и в принципе закон возрастания энтропии мог бы быть ее макроскопическим выражением... Однако до настоящего времени никому не удалось сколько-нибудь убедительным образом проследить такую связь и показать, что она действительно имеет место.

Вопрос о физических основаниях закона монотонного возрастания энтропии остается, таким образом, открытым. Не имеет ли его происхождение классической природы и не связано ли оно с общей проблемой начальных условий в космологии? Играет ли, и какую роль, в этом вопросе нарушение временной симметрии в некоторых процессах слабых взаимодействий между элементарными частицами? Возможно, что на подобные вопросы будут получены ответы лишь в процессе дальнейшего синтеза физических теорий» [8].

Более того, традиция связывать физические причины, приводящие ко второму началу термодинамики, с космологией восходит к Людвигу Больцману. Классик связывал возникновение звезд, планет, галактик с гигантским самопроизвольным отклонением от равновесия.

На новом уровне развития физической теории космологическую трактовку второго начала обосновывает и Р. Пенроуз: «Таким образом, напрашивается вывод, что та удивительно низкая энтропия, которую мы обнаруживаем вокруг себя — и которая составляет наиболее загадочную сторону второго начала термодинамики, — должна быть приписана тому, что огромный выигрыш в энтропии может быть получен в процессе гравитационного сжатия рассеянного газа в звезды... Именно возможность этого газа собираться в гравитационные сгустки и дала нам второе начало термодинамики».

В отличие от Больцмана, высказывавшего смелые догадки, Пенроуз может опираться на основательный фундамент общей теории относительности и релятивистской астрофизики, ведущим специалистом в которых он является.

В книге сообщается еще более удивительный факт. Начальные условия, которые могли привести к возникновению нашей Вселенной в результате Большого взрыва, также должны были быть удивительными. Объем той части фазового пространства возможных вселенных, в котором возникает реальность, близкая к нашей, с «правильным» вторым началом термодинамики, составляет всего $1/10^{10^{123}}$.

Это число настолько мало, что все сравнения тут бледнеют. В соответствии с современными физическими воззрениями, атомов во Вселенной существует около 10^{80} (столько же банковских счетов может быть занумеровано с помощью 80-знакового числового кода, который лет 5 назад был принят в банковской системе России). Один выдающийся математик обратил внимание на то, что в математических исследованиях не фигурировали числа, большие 10^{100}, и присвоил последнему название «гугол». Имеют ли физический смысл такие большие числа, как «гугол», и такие малые, как «обратный гугол»? Насколько велико должно быть «бесконечно большое» или мало «бесконечно малое»?

Вопрос старый. Помнится, еще Беркли иронизировал над введенными Ньютоном «бесконечно малыми», называя их «тенями безвременно усопших величин». В каждой физической или математической теории, претендующей на какие-то отношения с реальностью, приходится оценивать, что разумно считать малым, а что большим.

Но, очевидно, то же касается и фундаментальных теорий, определяющих наше мировоззрение. Интересной и глубокой представляется радикальная точка зрения, недавно выдвинутая Д. С. Чернавским [9]: «Физической причиной возрастания энтропии и необратимости процессов во времени является неустойчивость динамических систем. (Как классических, так и квантовых.) Однако для конкретного математического описания неустойчивых процессов необходимо дополнить математическую аксиоматику утверждением:

Корреляции высокого порядка между случайными величинами (*порядка „обратный гугол", т. е.* 10^{-100}, *и выше*) *должны быть признаны отсутствующими*, даже если они возникают в аналитических расчетах.

Основанием для этого можно считать следующее.

Во-первых, такие корреляции физически не реализуемы, т. е. их в принципе невозможно ни наблюдать, ни проверить.

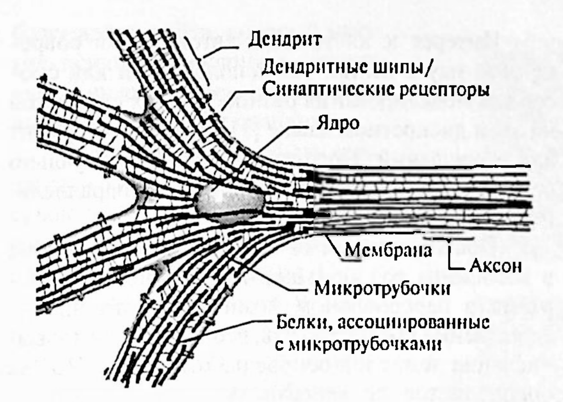


Рис. 1

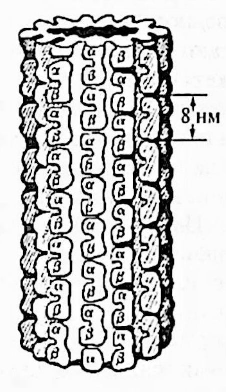


Рис. 2

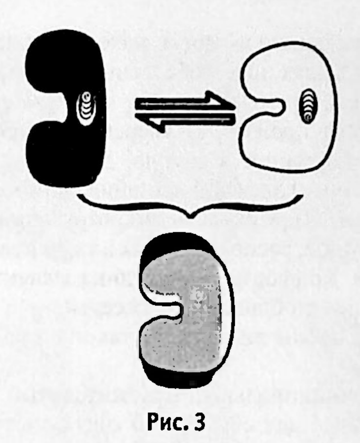


Рис. 3

Во-вторых, в этом и только в этом случае расчеты ведут к наблюдаемым результатам, т. е. описывают необратимые во времени процессы.

В-третьих, это правило уже давно используется на интуитивном уровне при решении конкретных задач».

Естественно ожидать, что похожим образом дело обстоит и с фазовым пространством возможных вселенных. Число $10^{-10^{123}}$ говорит само за себя. И сталкиваясь с настолько невероятным событием, которое тем не менее произошло, естественно предположить, как в хорошем детективе, что мы не знаем чего-то очень существенного.

Завораживает логика Р. Пенроуза. Она примерно такова. Интуитивно ясно, что мозг должен чем-то радикально отличаться от компьютера. Те есть решать задачи, которые компьютеру не под силу, но с которыми человек справляется. Современная математика убедительно показывает, что таковые есть. Значит, «физика мозга» отличается от принципов работы компьютера.

Но поскольку в обычной, хорошо понятой физике таких отличий не видно, то стоит поискать ответ на границе между незнанием и нашим сегодняшним знанием. Граница же проходит там, где лежат «трудные вопросы» — о механизме редукции волнового пакета, о стреле времени, о согласовании теории гравитации с квантово-механическими представлениями. А теперь предположим — следите за ходом мысли автора (или логикой Природы?) — что все эти задачи имеют одну общую отгадку. Пенроуз называет ее ПКТГ — *правильной теорией квантовой гравитации*. Все собирается воедино. Появляется простота. Ясность и единство.

Итак, новый наряд короля пока не сшит (и неясно, можно ли его сшить вообще), но скроен Р. Пенроузом он виртуозно!

Первые успехи

А какая польза от всего этого будет нашему колхозу?

С. Э. Шноль

Успех исследовательской программы определяется не только авторитетом ее автора и парадоксальностью предложенной концепции. Во многом он зависит от научных результатов других исследователей, начавших разработку этой программы. И такие результаты уже есть.

Один из главных состоит в обнаружении конкретных структур в нервной клетке — нейроне, — которые могли бы отвечать за «невычислимость сознания», связанную с объективной редукцией.

Исследователь из Аризонского университета в Тусоне Стюарт Хамерофф (Stuart Hameroff) в 1997 году выпустил книгу “Ultimate Computing: Bimolecular Consciousness and Nano Technology” [10]. Речь в ней идет о возможности своеобразных вычислений, происходящих в *микротрубочках цитоскелета* (рис. 1). Эти объекты представляют собой важную часть «скелета» клетки (рис. 2), они имеют диаметр примерно — 25 нм. И состоят из молекул — димеров *тубулинов* (рис. 3).

Молекулы димеров могут находиться по крайней мере в двух пространственных конфигурациях или, как говорят биологи, *конформациях*. Для того чтобы произошло «переключение» из одной конформации в другую, достаточно, чтобы единственный электрон изменил свое положение в молекуле. Поверхность микротрубочки состоит из тубулинов, расположенных в узлах правильной решетки. Конформация тубулина зависит от конформаций его ближайших соседей.

Способны ли системы такого типа к вычислениям?

Принципиальный положительный ответ на этот вопрос дал еще в 1970 году кембриджский математик Джон Конвей. Он придумал для своих студентов компьютерную игру «Жизнь», представляющую собой так называемый *клеточный автомат*. Клеточный автомат моделирует процессы, разворачивающиеся в пространстве и во времени. Пространственные и временная координаты в таких моделях дискретны, а сама ячейка-клетка может находиться в одном из конечного набора состояний. Кроме того, состояние каждой клетки на следующем шаге по времени зависит от состояния ее ближайших соседей в данный момент.

Игра «Жизнь», к примеру, представляет собой автомат, заданный на бесконечном листе в клетку. Сами клетки могут находиться в двух состояниях — «живом» и «мертвом». «Живая» в данный момент клетка в следующий момент становится «мертвой», если у нее менее двух или более трех «живых» соседей. «Мертвая» клетка «оживает», если у нее ровно три «живых» соседа. Осталось сказать, что «соседями» данной клетки являются 8 других, которые имеют с ней общие ребра или вершины.

Эта игра должна была имитировать рост и развитие колоний микроорганизмов. Джон Конвей показал, что, несмотря на поразительную простоту правил «Жизни», этот клеточный автомат эквивалентен машине Тьюринга, то есть способен выполнять вычисления любой сложности.

Игра «Жизнь» не очень похожа на то, что можно ожидать на квантовом уровне. Она относится к так называемым *диссипативным* системам, где нет сохраняющихся величин и детали начальной конфигурации могут быть забыты. В фундаментальных физических процессах, в реальных процессах на наномасштабах дело обстоит несколько иначе. В них сохраняется энергия и система «не умеет» забывать начальное состояние. Такие объекты называют *гамильтоновыми* системами. Тем не менее выдающийся физик-теоретик XX века Ричард Фейнман показал, что и в гамильтоновых клеточных автоматах также могут производиться сколь угодно сложные вычисления.

Интерес к клеточным автоматам в современной науке велик. Их используют и как способ для моделирования разнообразных процессов на этом дискретном языке [11], и как инструмент для вычислений. Последнее особенно актуально сейчас, когда многие задачи требуют параллельных вычислений.

Поскольку «время срабатывания» нейрона в миллионы раз меньше, чем логического элемента в персональном компьютере, то вполне естественно предположить, что параллельные вычисления лежат и в основе работы мозга. Но ряд специалистов по *нанобиологии*, основываясь на идеях Пенроуза и Хамероффа, косвенных экспериментальных свидетельствах и физических соображениях, предполагают большее.

Они считают, что большая совокупность тубулинов может некоторое время развиваться в полном соответствии с законами квантовой механики, а затем претерпевать объективную редукцию (несмотря на то, что наблюдателя или процедуры наблюдения нет!) и переходить в классическое состояние. Именно эта редукция и отвечает за момент осознания.

Сказанное иллюстрируют рис. 4 и рис. 5. На них черные и белые «крючки» — тубулины в разных конформациях. Серые «крючки» — совокупность тубулинов, находящихся в квантовой когерентной суперпозиции. На рисунках показан рост этой совокупности вплоть до момента объективной редукции, после которого возникает классическое состояние автомата.

Обратим внимание на то, что все это коррелирует с понятием *инсайта* — одного из фундаментальных понятий современной психологии. Это понятие (от английского insight — проницательность, проникновение в суть) было введено в 1925 году В. Келлером. Оно означает внезапное усмотрение сути проблемной ситуации. В опытах Келлера с человекообразными обезьянами животным предлагались задачи, которые могли быть решены только опосредованно. Исследование показало, что обезьяны после ряда безрезультатных попыток прекращали активные действия и просто разглядывали предметы вокруг, после чего достаточно быстро приходили к правильному решению. В дальнейшем это понятие использовалось К. Дункером и М. Вертгеймером в качестве характеристики человеческого мышления, при котором решение достигается путем мысленного постижения целого.

Исследование целостного восприятия образа, проблем ситуации стало основой большого научного направления — гештальтпсихологии. Синергетика связывает целостное восприятие с отысканием сознанием параметров порядка в исследуемом объекте. Концепция Пенроуза—Хамероффа предлагает возможную «элементную

базу» для инсайта, который многие психологи склонны считать «некомпьютерным», «невычислимым» элементом сознания.

Исследователями были сделаны и количественные оценки совокупности тубулинов, которые могут находиться в сцепленном квантовом состоянии. С одной стороны, Пенроузом была предложена формула, связывающая продолжительность «квантовой жизни» с гравитационной энергией системы. С другой стороны, существует ряд экспериментов в нейрофизиологии, которые позволяют оценить характерные для нервной системы величины. Эти оценки показывают, что для реализации «момента сознания» нужно примерно 1 миллиард тубулинов. Для этого вполне достаточно 1000 нейронов. Количество таких событий может достигать сотен миллионов в секунду. (В этом контексте инсайт, над объяснением которого бьются психологи, — сущность более высокого порядка.) Естественно, «моменты сознания» будут в соответствии [10] с этой гипотезой иметь место и у других видов. В частности, в нервной системе червя с 302 нейронами и примерно тремя миллиардами тубулинов моменты сознания будут происходить не чаще двух раз в секунду.

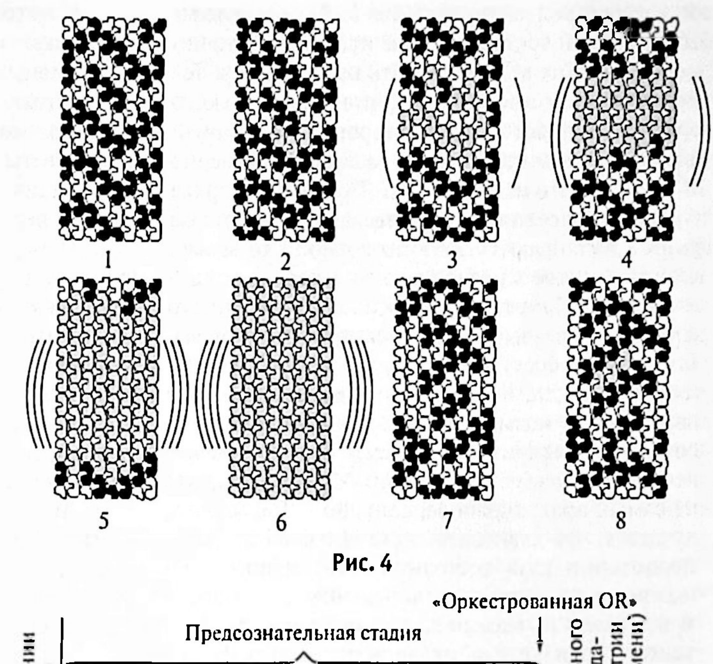


Рис. 4

«Оркестрованная OR»
Предсознательная стадия
Число тубулинов в квантовом когерентном состоянии
n
Проявление квантовой когерентной суперпозиции
Классические вычисления
Квантовые вычисления (унитарная эволюция квантовой системы)
Момент сознательного опыта (выбор фундаментальной геометрии пространства-времени)
0
500
Время, мс

Рис. 5

Итак, развитие концепции Р. Пенроуза в контексте поиска кандидатов на роль биологических структур, в которых могла бы происходить объективная редукция, доведена до уровня оценок и конкретных экспериментальных постановок.

До этого уровня доведена и другая идея, связанная с объективной редукцией. Р. Пенроуз с коллегами предложил эксперимент, который может стать ключевым в «правильной теории квантовой гравитации», о которой Р. Пенроуз рассуждает в книге «Новый ум короля». По существу, это вариант знаменитой «шредингеровской кошки». В обоих случаях квантовый объект оказывается связанным с классическим, и если принципы квантовой механики верны, то он должен оказываться в суперпозиции двух состояний (то есть кошка должна быть одновременно и чуть-чуть жива, и чуть-чуть мертва). Схема эксперимента показана на рис. 6.

Суть этого эксперимента Р. Пенроуз описывает следующим образом [10, 12]:

«Итак, слева находится источник фотонов, за ним — расщепитель луча, который создает два луча в квантовой суперпозиции. Внизу — маленькое зеркало, оно находится на кантилевере (это микроскопическая пружинка). Когда фотон попадает в зеркало, он заставляет его вместе с пружинкой слегка перемещаться, и зеркало чуть-чуть раскачивается. Дальше, вот



Рис. 6

здесь ставятся два зеркала, a и b. Фотон должен быть рентгеновский, так как нужно достаточно энергии, чтобы воздействовать на пружинку. Теперь система, состоящая из фотона и маленького зеркала, начинает эволюционировать как единый квантовый объект. Зеркало содержит примерно 10^{13} атомов, его масса 10^{-13} кг. Положение зеркал и размах колебаний кантилевера под действием фотона подбираются так, что ровно за то время, пока маленькое зеркало на кантилевере совершает одно колебание, луч отражается от удаленного зеркала b и возвращается к маленькому зеркалу. Там два луча соединяются и идут наверх к сплиттеру, и т. д. Да, я не сказал, что вся установка находится в космосе, и расстояния между некоторыми элементами системы будут не менее нескольких диаметров Земли. Может быть, это не самый практичный вариант, но... Так вот, получается, что квантовая система находится в суперпозиции двух состояний, отвечающих двум положениям фотона — попаданию в зеркало a и в зеркало b. Наконец, вверху помещается детектор. Если этот детектор регистрирует фотон, значит суперпозиция системы разрушилась. Есть очень серьезные, на мой взгляд, теоретические аргументы в пользу того, что эта суперпозиция, даже при условии тщательной изоляции от окружающей среды (для чего и нужна орбитальная платформа — вакуум, сверхнизкие температуры), должна спонтанно коллапсировать в одно из двух возможных состояний, причем в течение вполне определенного интервала времени, который вычисляется на основе гравитационных соображений».

По этому интервалу времени τ определяется *собственная гравитационная энергия*

$$E_G \approx \tau h,$$

где h — постоянная Планка.

Само существование этой величины и нетривиального интервала τ будет означать правоту Р. Пенроуза, а значение E_G позволит судить о точности его теории.

Эксперимент не прост — нужны зеркала фантастического качества — более миллиона отражений без существенной потери энергии. Но одни энтузиасты планируют такое на орбите, а другие ищут альтернативы космическим экспериментам [13].

Сейчас в фундаментальной науке очень часто появляются теории, для проверки которых надо либо подождать еще 600 миллионов лет, либо заглянуть на масштабы в 10^{-35} см, или посмотреть на происходящее через время 10^{-20} с после Большого взрыва, либо, на худой конец, получить в свое распоряжение десяток черных дыр или пяток вселенных, устроенных несколько иначе, чем наша.

И в этом контексте подход Р. Пенроуза представляется исключительно интересным и конструктивным. В концепции Пенроуза задаются конкретные вопросы, на которые в ближайшем будущем могут быть получены экспериментальные ответы. И в случае отрицательных ответов концепция должна быть отброшена либо радикально перестроена.

Первые успехи подхода, выдвинутого Р. Пенроузом, позволяют высказать несколько методических замечаний. Если эксперименты дадут положительный ответ — победителей не судят, и замечания только покажут, какой сложный выбор пришлось сделать исследователю и какую блестящую интуицию проявить. А если нет... То на вопросы, поставленные в «Новом уме короля» придется искать другие ответы.

Вначале о «невычислимости» в нанобиологии. Допустим, что Р. Пенроуз прав, и «элементная база» сознания, в отличие от взгляда стандартной теории, имеет дело с невычислимыми сущностями. (В теории нейронных сетей, которую сейчас рассматривают как основу нейронауки, все вычислимо — нейрон представляется как сумматор с различными весами суммирующих входные сигналы и нелинейный элемент, вырабатывающий стандартный сигнал, когда величина суммы превысит некоторый порог.) Но это «микрооснова» сознания. А математическое творчество, сочинение музыки, о которых пишет Р. Пенроуз, инсайты, наконец, это макрособытия. И путь от «микроуровня» сознания и психики к «макроуровню» может быть очень длинным.

В самом деле, чтобы невычислимость с микроуровня «поднялась» на макроуровень и обусловила такие замечательные процессы, как математическое творчество, требуются механизмы гигантского и избирательного усиления. Может быть, именно с ними и связана самая трудная и интересная часть проблемы. И тогда, вероятно, не очень важно, что является «сырьем» для усиления и отбора — результаты работы множества «вычислимых» логических устройств, или множества микросистем, использующих объективную редукцию или даже какую-то часть шума.

Информация — это случайный запомненный выбор. А чем обеспечивается эта случайность — вопрос второстепенный. Гораздо интереснее во многих задачах понять, как запоминается необходимое, как редактируется «поток возможностей», в котором мы находимся, как появляется запрос на дополнительную или корректирующую информацию.

В том, что касается творчества, также следует признать пробел в наших знаниях. Мы очень мало знаем об интуиции. Мало, в сравнении с той ролью, которую она играет в нашей реальности.

Можно ожидать, что в ближайшем будущем появятся концептуальные и математические модели, связанные с интуицией.

Кроме того, вероятно, в процессе творчества огромную роль играет метод проб и ошибок. Выбор в огромном пространстве возможностей, который опирается на интуитивное представление (например, о гармонии или красоте), на аналогии, на предшествующий опыт автора. И, возможно, именно на этом уровне лежит понимание процесса творчества и связанные с этим «макромодели».

С гигантским усилением неразрывно связана и основа всей концепции Пенроуза — редукция. В самом деле, что такое «наблюдатель»? Это устройство, которое может усилить квантовый эффект и перевести его на макроуровень. Если бы мы, следуя главной идее Пенроуза — искать основу сознания на квантовом уровне, — могли бы увидеть этот механизм гигантского усиления квантовых эффектов в нервной системе, то гипотезы об объективной редукции просто не понадобились бы.

В самом деле, давайте взглянем со стороны на предлагаемую логику исследования. Замечательный и непонятный феномен сознания мы хотим объяснить гипотетическим и пока не обнаруженным феноменом объективной редукции, который, возможно, будет следовать из пока не созданной «правильной теории квантовой гравитации». Дерзкий и захватывающий полет мысли.

Но если мы готовы играть на такие ставки и делать другие настолько смелые допущения, то, наверно, стоит взглянуть и на альтернативы. А они есть.

В самом деле, говоря о редукции волнового пакета и выдвигая гипотезу об объективной редукции, мы становимся похожи на нерадивого студента. Вместо того чтобы по существу ответить на вопрос, с чем связано наблюдаемое и принципиально важное явление субъективной редукции, мы придумываем что-то совсем новое, о чем ни мы, ни экзаменаторы не имеем ясных представлений (впрочем, некоторые экзаменаторы ценят такой полет мысли).

В чем парадоксальность процесса редукции? В том, что одна частица, которая прекрасно описывалась уравнением Шредингера, ψ-функцией вдруг при наблюдении оказывается локализованной *в одном месте*. Физически проблема состоит в том, что стандартная квантовая механика никак *не объясняет* поведение квантовых объектов, а лишь *предсказывает* амплитуды вероятностей элементарных процессов.

Интерес к основам квантовой механики, связанный с квантовой информацией, квантовыми компьютерами и шифрами, возник в последнее десятилетие. Он во многом связан с мысленным экспериментом Эйнштейна, Подольского и Розена, предложенным в 1935 году (ЭПР).

Об этом эксперименте и его практическом воплощении прекрасно пишет Р. Пенроуз. Суть ЭПР эксперимента состоит в следующем. Квантовая система разделяется на две подсистемы в таком процессе, где сохраняется некоторая *суммарная* величина, например, общий угловой момент (спин). Подсистемы удаляются на произвольно большое расстояние. Если после этого измерить сохраняющуюся величину для одной подсистемы, то результат такого же измерения для второй системы становится однозначным. Стандартная квантовая теория не объясняет, почему и *как* спины произвольно удаленных частиц оказываются коррелированными. Более того, в принципе нельзя утверждать, что коррелированные величины «существуют» непосредственно после разлета частей системы, «просто мы их не знаем». Проблема состоит в том, что каждая из подсистем принимает свое значение данной величины только в *момент* ее детектирования измерительным прибором, И это еще один парадокс редукции.

С целью *объяснить* такие нелокальные корреляции и, по существу, с целью *явно описать* процесс перехода квантовой системы из одного состояния в другие, А. Эйнштейн ввел гипотезу «скрытых параметров». Непрерывная и детерминированная эволюция скрытых параметров должна была, по его мысли, давать «на выходе» наблюдаемые значения физических величин. Работы Белла и соответствующие эксперименты заставили отказаться от идеи скрытых параметров.

Однако появляются новые подходы к старым проблемам. Обращу в этой связи внимание на *концепцию скрытого времени*, выдвинутую П. В. Куракиными дающую новый взгляд на проблему редукции [14].

Суть этой концепции такова. Следуя логике квантовой механики, надо определить процедуру измерения физического времени. Традиционный взгляд неявно использует интуитивное классическое допущение о равномерном течении времени как некого абстрактного потока, на фоне которого разворачиваются все события. Но это противоречит квантовой природе элементарных событий.

В самом деле, чем измеряется время? В специальной теории относительности А. Эйнштейн ввел понятие «часов», но четко его не определил. Насколько мелкие доли «поворота» стрелки часов мы можем различить? Ответ П. В. Куракина, с которым трудно не согласиться, таков:

«Поскольку каждый, самый маленький поворот „стрелки“ часов сопровождается поглощением энергии, то ответ на последний вопрос ста-

(D2) ←---------------- (S) -----------→ (D1)

(D2) ----------------→ (S) ←----------- (D1)

(D2) ←•••••••••••••••• (S) -----------→ (D1)

Рис. 7

новится очевидным. Энергия квантуется, значит, квантуется и измерение времени. Таким образом, в каждом конкретном эксперименте время измеряется количеством поглощенных квантов энергии.

С точностью до нормирующего множителя, обусловленного конфигурацией конкретного эксперимента, время прошедшее в данной точке, *есть* количество поглощенных квантов энергии в данной точке.

Важно признать, что количество поглощенных квантов не просто *один возможный способ* измерить время, это в принципе *единственный способ*. Все остальные сводятся к этому».

Но тогда, следуя традиции Эйнштейна, мы можем ввести в теорию набор переменных, которые не являются физически наблюдаемыми величинами, а входят только в *математический аппарат* теории. Эти переменные эволюционируют в так называемом «внутреннем времени теории», которое не тождественно физическому времени и является математическим понятием. При этом элементарные события (такие как поглощение фотона атомом) являются «точками сшивки» внутреннего времени и физического времени.

Но в свете этого предположения и редукция, и парадокс ЭПР объясняются элементарно.

Рассмотрим случай одного фотона и разберемся, почему он оказывается в конце концов зарегистрирован *одним* детектором, а не двумя сразу (т. е. почему здесь происходит редукция).

Принципиальное объяснение иллюстрирует рис. 7. Излучение фотона источником и поглощение его детектором происходит в результате трех проходов сигнала скрытых переменных от источника к детектору и обратно:

- сначала источник посылает так называемые *сигналы поиска* по всем направлениям в пространстве во внутреннем времени;
- все потенциальные детекторы, получившие *сигнал поиска*, посылают источнику свои сигналы запроса на получение фотона;
- только один атом-детектор получает *сигнал подтверждения*, остальные получают сигнал *отказа*.

И, наконец, фотон оказывается в детекторе, получившем сигнал подтверждения.

Есть веские основания считать пространство дискретным — своеобразной решеткой с ребрами и узлами. При увеличении *сигналов запроса* (они движутся строго по тем же ребрам, что и сигналы поиска, но в обратном направлении) от потенциальных детекторов они конкурируют в каждом узле. Правила конкуренции просты: в каждом узле выживает ровно один входящий сигнал запроса, после чего он копируется во все ребра с входящим сигналом поиска. Это правило гарантирует, что в итоге останется только ровно одна «дорожка» от источника к единственному детектору. Этот детектор и получает «приз» — регистрирует испущенный фотон.

Другими словами, выбор осуществляет все пространство, вся Вселенная, причем во внутреннем времени. И у такого взгляда есть давняя квантово-механическая традиция — в формулировке квантовой механики Р. Фейнмана амплитуда вероятности перехода частицы вычисляется суммированием некоторого интеграла вдоль всех возможных путей из начального состояния в конечное.

Цена за объяснение редукции и квантово-механических парадоксов велика. Но зато невычислимость «прилагается бесплатно». У нас нет алгоритма, показывающего каким будет *этот выбор*, и можно говорить лишь о статистике и т. д.

Поэтому может быть стоит посмотреть на *нанобиологию*, имея в виду не новую сущность — объективную редукцию, — а старую добрую субъективную редукцию.

И, пожалуй, еще одно методическое замечание. Наша реальность устроена так, что на разных масштабах в пространстве и времени влияние разных факторов меняется. Поэтому нам и удается выделить главное, пренебрегая второстепенным. Постоянная такой структуры, отвечающая за квантовый мир и электродинамику

$$\alpha = \frac{1}{137}.$$

Аналогичная постоянная, характеризующая гравитацию $\approx 10^{-42}$. Они отличаются на много-много порядков.

Может быть, в первые мгновения после Большого взрыва (если таковой был, конечно) или внутри черной дыры или в других вселенных (существование которых всерьез рассматривается в ряде модных физических теорий) квантовые и гравитационные эффекты играли в определенных явлениях сравнимую роль. Но в нашей реальности... В каждом акте сознания...

Здесь потребуются очень серьезные доказательства. Но если в ходе дальнейших поисков устоит хотя бы часть концепции Пенроуза, это будет прекрасно. Это будет вдохновляющим примером для многих ученых и стимулом для других междисциплинарных проектов.

По направлению к нелинейности

В этом новом, «контрастном» Сумраке двигаться было куда проще. Я поднял свою тень легко, будто в обычном мире. И оказался там, где дозволено находиться лишь магам вне категорий.

С. Лукьяненко

Когда отплываешь от берега быстрой реки, то чувствуешь, как властно течение стремиться вынести на быстрину, на фарватер. На мой взгляд, похожее явление имеет место с потоком научных исследований. Обычно с годами и фарватер реки, и ее русло меняются. То же происходит и с наукой. И сейчас все больше фарватер естественно-научных исследований начинает определять нелинейная наука.

Наверно, можно выделить несколько различных уровней исследований.

На первом люди имеют дело с наблюдениями, с натурным или вычислительным экспериментом. Объясняют наблюдаемое, строят феноменологические модели.

На втором создают теории, предлагают обобщения, формулируют принципы, опираясь на которые можно объяснить и описать многое. Именно таковы «превосходные» по терминологии Пенроуза, теории.

Но есть и третий уровень. Наверно, осознание того, что он есть, пришло во второй половине XX века. На нем лежат качественные представления о тех явлениях, которые описывают теории, множество простейших «очевидных» моделей, представления о том, какими должны быть теории, и «строительный материал» для создания таковых. Иными словами, на этом уровне находится «язык», на котором говорит современное естествознание. И все чаще в последние десятилетия на этом уровне оказываются идеи, методы, понятия нелинейной науки, синергетики.

Об этом убедительно свидетельствует и книга Р. Пенроуза. Очень приятно читать восхищенные строки автора, посвященные красоте фракталов, множеству Мандельброта, одной из базовых моделей нелинейной динамики — отображению $z \to z^2 + c$. Чудесно читать про путешествие в некий далекий мир «Тор'Блед-Нам» и про то, что каким-то понятиям нелинейной науки уже отведено место в платоновском мире идеальных сущностей. Более того, ключевые для книги понятия невыполнимости и нерекурсивности прекрасно иллюстрируются с помощью множества Мандельброта.

Не обойден вниманием и динамический хаос, чувствительность к начальным данным и горизонт прогноза. Однако приговор автора этим механизмам, с точки зрения сознания, достаточно суров. Этот момент принципиален, поэтому приведу довольно длинную цитату о динамическом хаосе, отражающую взгляд Р. Пенроуза. «Все это не похоже на тот тип „невычислимости“, который сложно было бы каким-то образом „использовать“. Невычислимость в данном случае обусловлена просто тем, что из-за существования предела точности, с которой может быть известно начальное состояние, будущее состояние в принципе не поддается точному расчету на основании известных начальных условий. На самом деле в этом случае к будущему поведению системы примешивается *случайный элемент* и только. Если же работа мозга все-таки опирается на *полезные* невычислимые составляющие физических законов, то последние должны быть совершенно другие — и более конструктивные по своей природе. Поэтому я не буду называть „хаотическое“ поведение такого рода „невычислимостью“, предпочитая использовать термины „непредсказуемость“. Наличие непредсказуемости — весьма общее явление для тех детерминистических законов, которые, как мы вскоре убедимся, действительно возникают в классической физике. Но мы скорее уж предпочтем минимизировать непредсказуемость, чем „использовать“ ее в конструкции мыслящих машин!»

Да нет же! Совсем наоборот, — так и хочется воскликнуть в этом месте. Думаю, что сейчас большинство «синергетиков» думают иначе. Коротко суть их возражений и идей, сводится примерно к следующему.

Несущественно, что «мы предпочтем», поскольку речь идет не о конструировании «мыслящей машины», а о том, как устроены природа и сознание. Одно из фундаментальных свойств мира — нелинейность и связанная с ней неустойчивость. Поэтому вполне естественно предположить, что «сознание», в первую очередь предназначенное для того, чтобы отражать нашу реальность, использует различные нелинейные механизмы и связанные с ними неустойчивости.

«Полезную», «ценную» макроскопическую информацию естественно рассматривать как *случайный запомненный выбор*. Простейшие модели нелинейной динамики (например, $x_{n+1} = 1 - 2|x_n|$) показывают как может генерироваться информация. Благодаря нелинейности и чувствительности к начальным данным несущественное усиливается и становится значимым. Не менее важен и другой нелинейный эффект.

Усиленный сигнал под влиянием внешних воздействий (то самое отражение, которым должно заниматься сознание) или случайно может быть запомнен! [9], поэтому «полезность» этого сорта невычислимости, детально исследованного в синергетике, не вызывает сомнений.

Проблема состоит в том, как сознание (или «мыслящая машина» следующего поколения) ориентируется в многомерном, многофакторном мире. Главное — алгоритм редактирования поступающей информации, и *забывания*, и прогноза (опережающего отражения, как говорят психологи). Чтобы понять это, была предложена теория *русел и джокеров* [15]. *Русла* — области небольшой размерности в фазовом пространстве с большим горизонтом прогноза. Благодаря наличию русел и нашей способности их выделять, мы и можем ориентироваться в мире, предвидеть последствия своих действий. Но в фазовом пространстве есть и *области джокеров*. Где горизонт прогноза мал, возможно быстрое изменение состояния системы. Можно сказать, что в областях джокеров может происходить гигантское усиление входных воздействий. Роль таких «гигантских усилителей» в природе, обществе, сознании, на мой взгляд, недооценивается. Мы не замечаем очевидного. Именно благодаря таким механизмам усиления и осуществляется связь между разными уровнями организации общества, организма, материи. Именно благодаря им мы что-то знаем о микромире (стандартный фотоумножитель способен «обменять» фотон на миллион электронов). Саму редукцию волнового пакета, взаимодействие микрообъекта с макроприбором, приводящее к изменению состояния последнего, также можно рассматривать как фундаментальную схему усиления. И если благодаря предвидению Р. Пенроуза удастся найти еще одну такую схему — объективную редукцию, то естественно было бы понять, используются ли сознанием эти механизмы и каким образом.

Зато другой элемент нелинейной науки — *сингулярности* (особенности) — занимает у Р. Пенроуза царское место. Фундаментальный вопрос, непосредственно связанный с мировоззрением — это существование стрелы времени, коренных различий между прошлым и будущим. Ответ на него можно искать как на микро-, так и на макроуровне.

Первый подход активно развивал один из отцов-основателей синергетики, Нобелевский лауреат Илья Романович Пригожин. По его концепции, необратимость самым тесным образом связана с неустойчивостью на микроуровне [16]. Наблюдатель и редукция волнового пакета при таком подходе ни при чем. Но квантовую механику надо несколько переформулировать, вводя диссипативные процессы и затухание на микроуровне. Математически это означает, что нам придется иметь дело с несамосопряженными операторами, более сложными функциональными пространствами. Соответствующий математический аппарат развивался И. Р. Пригожиным и его учениками совместно с сотрудниками МГУ им. М. В. Ломоносова в течение ряда лет.

В трактовке Р. Пенроуза причины появления стрелы времени, необратимости, второго начала термодинамики лежат в космологии, в удивительном типе особенности, которым началось существование Вселенной.

Сингулярностью или особенностью математики называют те точки, в которых функция или какие-то ее производные обращаются в бесконечность. Особенно любопытными представляются сингулярности, возникающие по мере того, как процесс разворачивается во времени. В этом случае говорят о развитии процесса (или росте решения) *в режиме с обострением*.

Очень любопытна история анализа таких замечательных процессов — любимого детища нелинейной науки — и изменения взглядов на этот объект.

С самого начала серьезного исследования нелинейных явлений с помощью компьютеров со средины XX века оказалось, что нелинейный мир полон режимов с обострениями.

Самый простой пример — ударные волны. Они возникают уже в простейшей одномерной версии гидродинамики — нелинейном волновом уравнении

$$u_t + uu_x = 0.$$

Это уравнение описывает волну, профиль которой становится все круче и круче, и в какой-то момент производная u_x обращается в бесконечность. И тогда надо или учитывать другие процессы (здесь — вязкость), или менять само понятие решения.

Подробнейшим образом изучены режимы с обострением в модели тепловых структур в научной школе член-корр. РАН С. П. Курдюмова

$$T_t = (K(T)\, T_x)_x + Q(T).$$

Здесь за конечное время температура $T(x,t)$ возрастает до бесконечности в одной точке или в целой области пространства.

Вначале на такие решения смотрели как на математический курьез, лишенный физического смысла. Но время шло, и такие решения возникали вновь и вновь в самых разных областях — от физики плазмы и проектов управляемого термоядерного синтеза [17] до демографии [3]. Поэтому к решениям, растущим в режиме с обострением, стали относиться как к полезному приближению (асимптотическому).

Но как только такое решение возникало, сразу становилась ясна ограниченность модели,

и исследователи «вынимали из рукава» какой-нибудь процесс, стабилизирующий рост функции или ее производных.

Поэтому очень интересно и важно, что и режимы с обострением, и сингулярность возникают при анализе «последних вопросов» — проблем начала и конца Вселенной. Роджер Пенроуз внес в это направление принципиальный вклад, показав, что для уравнений общей теории относительности, которые здесь приходится решать, такое поведение *правило*, а не исключение. Но так же, как в других задачах, связанных с режимами с обострением, это означает, что рассматриваемых физических процессов недостаточно, чтобы обеспечить стабилизацию (или, как иногда говорят, заглянуть за момент обострения). Нужны другие механизмы. Каковы же они? Это действительно глубокий мировоззренческий вопрос. Пенроуз возлагает надежды на ПТКГ — правильную теорию квантовой гравитации, которая может раскрыть эти механизмы. Может быть, будущее оправдает эти надежды.

Есть еще один важный «нелинейный» мотив в картине, создаваемый Пенроузом. И проблемы, связанные с вычислимостью, и многое другое иллюстрируется с помощью конечных автоматов, разнообразных машин, работающих с цифрами или символами. Но сами по себе они и есть нелинейные системы. Они основаны на наличии в природе пороговых явлений, дискретных сущностей. Давайте отдадим себе отчет в том, что мы являемся свидетелями и участниками «цифровой революции». Устройства, окружающие нас, все чаще работают не с аналоговыми непрерывными сигналами, а с цифрой. Заметим, что это одно из любопытных, а может быть и важных отличий мозга и компьютера.

На научных семинарах, проходивших в Институте прикладной математики им. М. В. Келдыша РАН, не раз обсуждались вопросы о дискретной формулировке законов природы. В самом деле, обратим внимание на непоследовательность и парадоксальность нашего нынешнего описания физической реальности. На микроуровне мир дискретен. Но законы природы в современной физике сформулированы на языке дифференциальных уравнений, оперирующих с гладкими и непрерывными функциями. Но поскольку в большинстве интересных и важных случаев мы без компьютера решать их не умеем, то возникает проблема «обратного перевода» — от непрерывных уравнений к тем дискретным сущностям, с которыми оперирует компьютер. Этот «перевод» сам по себе сложен и неоднозначен. Он породил многие направления современной математики, некоторые из которых особенно близки сердцу Р. Пенроуза.

И, естественно, возникает соблазн обойтись без перевода, формулируя законы природы не на «непрерывном», а на «дискретном» языке. Работы в этом направлении активно ведутся (здесь и машины клеточных автоматов, и теория самоорганизованной критичности, и другие подходы, дающие дискретную картину реальности). Тем не менее появление хотя бы одной «полезной» в смысле Пенроуза теории, сформулированной на дискретном языке могло бы в сущности повлиять на все естествознание.

На пути к постановке задачи

Читатель, вероятно, уже почувствовал, что у меня нет особенно убедительных аргументов в пользу своей собственной точки зрения. Если бы у меня были такие аргументы, я не стал бы так мучиться, разбирая ошибки, содержащиеся во мнениях противоположенных моему собственному.

А. Тьюринг.
«Может ли машина мыслить?»

Для того, чтобы работа стала классической и оказала большое положительное влияние на развитие науки, она не обязательно должна быть верной!

Из выступления на семинаре

Пойди туда — не знаю куда. Принеси то — не знаю что.

Из русской народной сказки

Часто бывает так, что ко времени сдачи отчета по научной работе становится ясно, как же надо было ставить задачу в самом начале и как организовывать работу. Надеюсь, что у некоторых читателей после прочтения книги возникает такое же ощущение — это дает надежду на прогресс в решении поднятых Пенроузом вопросов.

И чтобы таких читателей было больше, давайте попробуем вернуться к началу, к постановке задачи. Попробуем взглянуть на нее с точки зрения синергетики и нынешних компьютерных реальностей.

Обратимся к классической работе Тьюринга [18]. В ней общий и конкретный вопрос: «Мо-

жет ли машина мыслить?» предлагается заменить другим (потому что не определено, да и не очень-то ясно, что такое «машина» и что такое «мыслить»). А именно: «Может ли машина успешно играть в „имитацию"». То есть можем ли мы по ответам из-за стены выяснить, кто там находится — человек или вычислительная машина. Такова суть классического *теста Тьюринга.*

Эту же постановку задачи принимает и Пенроуз: «Несмотря на очевидную абсурдность некоторых аспектов рассматриваемого вопроса (в частности, моральных), мне кажутся достаточно обоснованными доводы в пользу того, что успешно пройденный тест Тьюринга *есть* указание на присутствие мысли, интеллекта, понимания или сознания. В самом деле, на чем еще могут основываться наши убеждения в присутствии этих качеств у других людей, кроме как на беседе с ними?.. Лично я готов к тому, чтобы значительно упростить тест Тьюринга... Мне бы хватило, чтобы наша проницательная опрашивающая по ответам на свои вопросы просто убедилась, что имеет дело с *сознательным разумом*, пусть даже чужеродным. Все то, что реально недостижимо во всех созданных по сей день компьютерных системах».

На мой взгляд, с этим сегодня нельзя согласиться. Алан Тьюринг писал более полувека назад. И тогда, естественно, многое из нашей сегодняшней и компьютерной реальности казалось чудом. В частности, «игра в имитацию». И написать программу, способную как поддерживать короткий разумный разговор на общие темы, — дело посильное для многих студентов факультетов компьютерных наук.

Можно вспомнить, к примеру, созданную в незапамятные времена программу «Элиза», имитирующую психотерапевта. Вспомним множество интерактивных игр и систем, успешно имитирующих наших человеческих партнеров. Более того, достижения современного искусственного интеллекта таковы, что системы поддержки принятия решений проходят тест Тьюринга, общаясь с профессионалами в их конкретных областях. Особо следует выделить системы, аккумулирующие, обобщающие и редактирующие опыт и знания высококвалифицированных врачей. Знание, как правило, невербализованное и неотрефлексированное [19]. По существу, речь идет о компьютерном профессиональном бессмертии. В шахматах — о компьютерных программах, которые обыгрывают мировых чемпионов.

Тьюринг, заглядывавший в компьютерное будущее и надеявшийся, что со временем компьютеры смогут проходить его тест, оказался прав! Давайте признаем это.

Но ведь и Роджер Пенроуз, и сторонники сильного искусственного интеллекта на самом деле говорят не о тесте Тьюринга. Они задают и обсуждают другой вопрос:

«Чем разум, мозг, сознание человека принципиально отличается от нынешних и будущих компьютерных имитаций?» (А)

По мнению сторонников сильного искусственного интеллекта, ничем. По мысли Роджера Пенроуза, отличается принципиально. Отличается, если отбросить оговорки, тем, что компьютерам принципиально недоступен ряд высших образцов математического творчества. В частности, они не могут решить несколько задач, с которыми блестяще справился Пенроуз и его коллеги. Очевидно, речь идет не о невинной болтовне с компьютером и «игре в имитацию». Планка установлена исключительно высоко!

Но прежде чем обсуждать утверждение (А) и уточнять противоречия, обратим внимание на точку соприкосновения и Пенроуза, и сторонников сильного искусственного интеллекта, и большинства специалистов по информатике.

Компьютеры, в принципе, могут справиться с любыми задачами, которые могут быть формализованы, которые имеют алгоритмы для своего решения и которые согласуются с физическими возможностями компьютеров. (Б)

В сущности, взгляд Пенроуза опирается на результаты о принципиальном наличии неразрешимых для компьютера задач и отсутствии способов вывести ряд утверждений в рамках формальных систем. Традиция противопоставлять «человеческую» и формальную логику характерна и для ряда психологических школ. В частности, ее развивал в течение ряда лет директор Института психологии, член-корр. РАН А. Н. Брушлинский.

Проще говоря, в соответствии с этим подходом, мы отличаемся от компьютера математической (и не только) интуицией и способностью выбирать и строить ту аксиоматику, которая лучше соответствует нашей реальности. И отсюда следует конструктивная постановка — построить компьютерные аналоги интуиции и научить правдоподобным рассуждениям — основе математического (и многих иных) видов творчества.

Вероятно, чтобы выявить отличие двух сущностей полезно сосредоточиться на главных, а не на второстепенных признаках. Согласимся, что немногие из нас доказывают такие прекрасные теоремы, как те, о которых пишет Пенроуз. И немногие из нас играют в шахматы на уровне чемпионов мира. И роль этих видов деятельности для нас тоже не слишком велика. Поэтому, ставя вопрос о теории сознания, глядя на нее с компьютерной точки зрения, важно было бы сосредоточиться на чем-то более фундаментальном. Обратим внимание на несколько возможностей.

Функциональный подход. Ощущение, что мы в каком-то высшем смысле выше компьютеров

(например, нам доступен мир идеальных платоновских сущностей или что-то еще), конечно, ласкает наше самолюбие. Но давайте встанем на материалистическую точку зрения (а будучи естественником трудно всерьез говорить о какой-то другой) и спросим себя, зачем нам интеллект. Ответ очевиден — для того, чтобы дать нам преимущество в ходе естественного отбора, помочь выжить, выделить наш вид среди множества других.

Для чего нужен искусственный интеллект или другие компьютерные имитации элементов мышления? Либо для того, чтобы выполнить наши поручения. (И в этом контексте сравнивать компьютерные программы и архитектуры с нашим сознанием так же нелепо, как руку человека с экскаватором. Странно равнять мастера с инструментом.) Либо использовать компьютер как средство узнать больше о нас самих. Иными словами, компьютер может здесь сыграть роль своеобразного «зеркала» для разных сторон нашего сознания. Простой пример. Давно уже не диво компьютерные программы, прекрасно играющие в шахматы. Они играют по-своему, по-компьютерному, оперируя со своими оценочными функциями, перебирая миллионы вариантов. И, конечно, интересно было бы использовать их. Чтобы понять, что же и как оценивает человек. То есть решить обратную задачу — постараться «увидеть», «смоделировать», понять способ нашего шахматного мышления. Естественно, тот же вопрос относится не только к мышлению, но и к восприятию, и разным видам творчества. Насколько мне известно, это направление, лежащие на стыке математики, психологии и компьютерных наук пока практически не развивается. Пожалуй, исключением являются работы М. А. Капустина из Московского физико-технического института, посвященные восприятию коротких мелодий и динамике освоения математических дисциплин. При этом модели нелинейной динамики здесь оказываются полезны [20]. Но, на мой взгляд, это «компьютерно–психологическое» направление заслуживает большего внимания.

Структурный подход. Выше мы говорили о программном обеспечении «сознания». Но, может быть, теорию стоит начать строить с анализа «аппаратной части»? Во многом именно таков пафос последних глав книги Р. Пенроуза.

Действительно, представим себе компьютер, в котором часть элементов постоянно выходит из строя. При этом мы не видим ни заметной утраты информации, ни «выпадения из сознания» освоенных алгоритмов. Это трудно вообразить, но наш мозг именно таков. Очевидно, было бы важно понять, какие архитектуры, способы кодирования записи и перезаписи информации могли бы обеспечить такое функционирование. Естественно, синергетика должна здесь сыграть ключевую роль. Это касается и самоорганизации нейронных ансамблей, и самоорганизации в информационном пространстве. Отчасти по этому пути идет современная теория нейронных сетей [21].

Искусственная жизнь. В основе этого направления, активно развиваемого в Институте сложности в Санта-Фе и в нескольких других научных центрах, жизнь рассматривается как *информационный феномен.* То есть организм рассматривается как информационная машина в отличие от традиционной биологии, трактующей его как химическую машину. Исследователи, развивающие это направление, считают, что «материальная основа» жизни несущественна. Это могут быть белковые тела, компьютерная программа, кремниевое создание или механический агрегат. При этом предполагается, что организм *полностью* может быть записан в виде некоторого текста, аналога компьютерной программы.

И надо признать, что Роджер Пенроуз в общем разделяет этот взгляд. Но он обращает внимание на одну тонкость: если рассматривать жизнь как квантовый феномен, если это принципиально важно, то надо «записать» состояние всех квантовых систем в данный момент. И тут возникают принципиальные трудности. Передав состояние одной квантовой системы другой квантовой системе, мы изменим состояние первой, что в принципе может необратимо испортить исходный объект. Так что любопытный вопрос «можно ли скопировать сознание?» и если можно, то «можно ли скопировать сознание, не разрушив исходный объект?» остаются открытыми.

И тут, пожалуй, стоит выразить сомнение относительно справедливости «компьютерной метафоры». В основе этой метафоры лежит предположение, что «сознание» — это некий «решатель задач» и можно четко разделить «программное обеспечение» деятельности мозга (которое и определяет сознание) и «аппаратную часть» (бренное тело, которое, в сущности, может быть любым).

Приведем несколько формальную аналогию из области математической физики. В самой основе этой дисциплины лежит возможность отделить главное — уравнения, описывающие физические процессы, — от второстепенного — начальных и граничных условий, отражающих конкретные детали. Однако при исследовании достаточно сложных или уникальных объектов (рождение Вселенной, эволюция биосферы, ряд экономических систем) такое разделение не проходит. В зависимости от начального состояния

объекта законы, по которым он будет развиваться, оказываются разными!

Естественно предположить, что сознание относится к такому же типу сущностей, что оно «не переносимо» с одного «компьютера» на другой.

В самом деле, то что нам особенно дорого в себе, за что мы себя любим и ценим — способность к логическому мышлению, рефлексии, язык и обучение, культура и многое другое — это вершина айсберга. О самом айсберге — работе нервной системы и деятельности психики мы знаем не очень много. А ведь именно «невидимая часть айсберга» и решает главные задачи, связанные с нашей жизнедеятельностью, — реагирование на внешние воздействия и координацию процессов в громадной химической машине. Мы представляем собой «химико-информационные» машины, и то, что первое можно отделить от второго, совсем не очевидно.

Вполне возможно, что здесь будут возникать парадоксы, схожие с «демоном Максвелла» в статистической физике. Решение «невинной» информационной задачи на этом уровне организации материи противоречит законам природы и означает существование вечного двигателя.

Наши представления о времени, пространстве, логике, не говоря уже о красоте, во многом «телесны» и предопределены нашим устройством. Поэтому каким будет «разум» или «сознание» существенно отличающихся от нас созданий, и по каким признакам можно судить о таковых — вопрос дискуссионный. Его с легкостью необычайной решают многие писатели-фантасты, но в науке, насколько я знаю, он всерьез не ставился.

Эволюционный подход. Вернемся к утверждению (А). По счастью, нам доступен еще один подход к решению этой задачи. Мы можем спросить себя, как возникало сознание, как появлялись и моделировались поведенческие стратегии, формы коллективного поведения, ритуалы и верования, как развивался язык. Это тем более удобно, поскольку все эти вопросы лежали на магистральном пути развития гуманитарных наук в XX веке. Другое дело, что здесь очень важен междисциплинарный подход, позволяющий сопоставить результаты наблюдений и анализа разным моделям сознания или построить последнее. В любом случае имеется большой материал, с которым интересно поработать.

Но сейчас появилась возможность подойти к проблеме и с другого конца, с «компьютерной стороны».

С одной стороны, и в математической психологии, и в теории нейронных сетей [15], параллельно развивались теории обучения. В отличие от традиционных исследований по искусственному интеллекту [22], где компьютеру сразу задавались программы действия в разных ситуациях, появилась новая возможность. Возможность построить системы, использующие самоорганизацию, возникновение адекватных реакций на внешние воздействия, появление взаимосвязей между различными стимулами.

Пока между тем, что знают, умеют и исследуют компьютерные науки и психоляогия — пропасть. Но если мы ее ясно видим, то можно подумать и о строительстве моста. И с мировоззренческой, и с прикладной точки зрения, дело того стоит.

В теории самоорганизации, синергетике активно развивается «парадигма сложности». В ее основе лежит парадоксальный результат, касающийся моделирования сложных систем. Например, если смоделировать биосферу с заданным набором биологических видов таким, как в настоящее время, то вся система быстро упрощается и вымирает. Однако если предположить, что виды меняются, приспосабливаясь друг к другу, то есть происходит *коэволюция*, то вполне возможно прийти и к нынешнему состоянию, и к другим разумным результатам.

Этот подход используется в научном направлении, развиваемом профессором В. Г. Редько [23] и называемом *эволюционной биокибернетикой*, для того чтобы описать усложнение биологических сообществ, возникновение поведенческих стратегий и в конечном счете логики. Полученные результаты очень интересны. Они позволяют увидеть возникновение «целей», «альтруизма», проследить разложение родового строя в примитивных сообществах.

Иными словами, «алгоритмы эволюции» сообщества достаточно быстро и эффективно определяют «логику» и «внутренний мир» индивида даже в сравнительно простых компьютерных моделях. Нельзя сказать, что это в полной мере согласуется с жесткой логикой сторонников сильного искусственного интеллекта. Скорее, это формализация «коллективного бессознательного», о котором уже век толкуют психологи.

Но это привносит в утверждение (А) совершенно новый эволюционный элемент. В самом деле, неразумно сравнивать мгновенный снимок объекта X с эволюционирующим миллионы лет объектом У. Разумно сравнивать «эволюцию X» с «эволюцией У» или, по крайней мере, ставить вопрос о коэволюции X и У. Например, подумать, хотим или готовы мы организовать эволюцию неких компьютерных сущностей. Любопытно в этой связи вспомнить стремительное развитие компьютерных вирусов, которые уже в состоянии приносить убытки в десятки миллиардов долларов в год. Вероятно, эволюци-

онный анализ утверждения (А) имеет большой смысл.

В этой связи напомню выражение Р. Пенроуза относительно эволюционного возникновения тех алгоритмов, которыми оснащено наше сознание: «Но я не понимаю, как естественный отбор сам по себе мог дать рождение алгоритмам, которые позволили бы делать осознанные выводы касательно *правомерности* применения всех прочих алгоритмов, которыми мы должны по идее пользоваться... Чтобы решить будет ли алгоритм действительно *работать*, нужно глубокое понимание, а не просто еще один алгоритм... трудно понять, как настоящие *усовершенствования* алгоритмов могут получаться таким вот случайным образом... Меня не покидает ощущение, что в самой эволюции, в ее явном „нащупывании" пути к какой-то будущей цели есть что-то загадочное и непостижимое».

Результаты исследований, о которых я упоминал, делают это выражение неосновательным. Видимо, анализ «глубокого понимания» и «алгоритмов отбора алгоритмов» в сфере логики — дело будущего. В том, что касается «случайности» в отборе, мутаций, то здесь, видимо, ситуация схожа с той, с которой сталкивались биофизики, анализируя биологическую эволюцию. Действительно, считая вероятности случайных мутаций, можно сделать вывод о невозможности появления сложных организмов. Но исследователи выяснили, что природа не меняет «букву» на «букву», а редактирует, как опытный редактор, вставляя или выбрасывая целые куски текста. При этом шансы получить что-то интересное значительно возрастают. Возможно с «алгоритмами синтеза алгоритмов» происходит что-то похожее.

Искусственный инстинкт. Сознание, понимающееся как способность к творчеству, является одной из высших и, вероятно, наиболее сложных форм отражения реальности, доступных живому. Поэтому естественно было бы начать с чего-либо более простого, с которым и может быть связан ключ к сложному вопросу.

В самом деле, в большинстве случаев от компьютерных программ требуется не «интеллект», а «инстинкт», — четкие стереотипные действия, возникающие в качестве ответа на определенный класс ситуаций. Само выражение «искусственный интеллект» принадлежит известному фантасту и футурологу Станиславу Лему.

Роджер Пенроуз осторожно говорит о сознании у животных. (И это естественно, поскольку он не определил этого термина и поднял планку слишком высоко.) Но, может быть, восхождение, опирающиеся на результаты современных физических теорий, к глубокому пониманию сознания, стоит начать с подножия. С инстинктов, с насекомых, с наиболее простых действий. Ведь если есть «невычислимость», с которой автор книги связывает сознание, если есть кандидаты на роль тех микроструктур, в которых должна была бы происходить объективная редукция, то было бы странно, если бы всем этим был наделен только человек.

Детективный подход. В истории науки порой оказывалось, что анализ какой-нибудь детали, несоответствия в стандартных теориях приводил к открытиям, менявшим парадигму. В начале XX века такими «несоответствиями» были излучение абсолютно черного тела и специальные вопросы электродинамики движущихся тел.

Напротив, другие парадоксальные несоответствия, от объяснения которых многого ждут, многие десятилетия обманывают ожидания. Например, удивительная способность молекул живого вращать плоскость поляризации проходящего света была открыта еще Л. Пастером. Однако огромные усилия, вложенные в ее исследования и объяснение, пока не дали того глубокого понимания феномена жизни, на которое надеялись ученые.

Так ли дело обстоит в психологии, нейронауке? И да, и нет.

Да, потому что несоответствий, удивительных особенностей, парадоксальных экспериментов, замечательных наблюдений много. Слишком много. У Роджера Пенроуза целая глава посвящена особенностям восприятия «скорости срабатывания» сознания. Мне с коллегами в течение ряда лет казалось удивительным «широкое использование» хаоса в нервной системе. Это, на первый взгляд, разительно отличается от «компьютерной метафоры» для сознания [15]. Была выдвинута правдоподобная гипотеза, объясняющая роль хаоса в мозге. Эти исследования получили интересное развитие в Канаде [21]. Более того, анализ количественных характеристик хаоса, порождаемого электрической активностью мозга, используется в ряде медицинских методик. Тем не менее должного внимания к этому подходу уделено, вероятно, не было. Потому что есть много других интересных альтернатив.

И тут можно дать вторую часть ответа на вопрос. «Детективный метод» хорошо и надежно срабатывает, когда непонятно немногое, когда понятно чему удивляться, когда есть стандартная теория. По-видимому, этот уровень в нейронауке пока не достигнут.

Другими словами, существует ряд серьезных и перспективных альтернатив подходу, выдвинутому Роджером Пенроузом. Тем не менее новизна и парадоксальность его идей заслуживает и восхищения, и внимания.

Продолжение следует

Эта модель покорила воображение и завладела мыслями целого поколения ученых. Она обеспечила основы для споров, породивших теории. Его модель оказалась столь полезной, что пробудила активность не только математиков, но и представителей ряда инженерных дисциплин. Привлеченные его доводы не всегда формальны, а последующие построения не всегда абстрактны.

А. Дж. Перлис.
Речь на присуждении
премии Тьюринга

Итак, Роджер Пенроуз предложил контуры естественно-научной картины мира. С большой энергией, упорством и настойчивостью он убеждает — у нас нет теории сознания. И когда теория появится, на какой бы основе она не начала строиться, этот пафос и эти усилия не будут забыты ни специалистами, ни историками науки.

Ему удалось очень удачно обострить противоречия, лежащие в основе нашего видения мира. Это и непонимание сути сознания, и загадка, связанная с редукцией волнового пакета. Это парадоксы, связанные с энтропией, космологией и стрелой времени. Наконец, это самая сильная и убедительная трактовка антропного принципа, связанная с общей теорией относительности и моделью Большого взрыва. В соответствии с ней мы живем в еще более «невозможном» мире, чем ученые думали раньше!

Сформулированы принципиальные проблемы, от которых нельзя отмахнуться. От ответов на них существенно зависит научная картина мира. Если надеждам Р. Пенроуза суждено оправдаться, то эта картина может приобрести стройность, единство и полноту, на которые ученые предыдущих поколений не смели и надеяться.

Подчеркну еще раз, что сам подход, развитый в книге, является плоть от плоти синергетическим (наверно, философы сказали бы здесь убедительные слова про постнеклассическую науку). В чем это проявляется?

В постановке общих вопросов, выходящих за рамки и отдельных теорий, и отдельных наук. В стремлении использовать идеи и опыт как естественных, так и гуманитарных дисциплин.

Во множестве моделей, образов, идей, пришедших из нелинейной науки. Здесь и фракталы, и динамический хаос, и множество Мандельброта, и клеточные автоматы, и качественный анализ динамических систем.

Наконец, это сам способ организации междисциплинарных научных исследований. Обращение к единомышленникам и оппонентам с помощью «книги нового поколения». Создание таким образом своеобразного «незримого колледжа» исследователей, работающих над междисциплинарным анализом сознания, над построением научной картины мира в предложенном Р. Пенроузом стиле.

Правильная, точная постановка проблемы — половина успеха. И синергетика помогла выполнить эту часть работы.

Но хочется надеяться на большее. И на то, что специалисты по синергетике активнее и успешнее будут браться за «вечные проблемы естествознания», за мировоззренческие вопросы. И на то, что специалисты в разных областях будут все чаще мыслить и говорить на языке синергетики, пусть даже не отдавая себе в этом отчета. И на то, что базовые нелинейные модели — основа синергетики — будут приходить не только из «классики» (гидродинамики, химии, физики плазмы, биологии), но и из «высших разделов науки» — общей теории относительности, квантовой механики, теории элементарных частиц, космологии и прочих.

Но главная надежда состоит в том, что синергетика поможет не только поставить и конкретизировать глубокие вопросы, затрагивающие наше мировоззрение, но и получить на них содержательные ответы. Поможет реализовать ту захватывающую исследовательскую программу, которая намечена в этой книге Р. Пенроузом.

Председатель редколлегии
Г. Г. Малинецкий

Литература

1. *Гарднер М.* Математические головоломки и развлечения / Пер. Ю. А. Данилова. М.: Мир, 1971.
2. *Данин Д.* Неизбежность странного мира. М.: Молодая гвардия, 1962.
3. *Капица С. П., Курдюмов С. П., Малинецкий Г. Г.* Синергетика и прогнозы будущего. 3-е изд. М.: УРСС, 2003. (Синергетика: от прошлого к будущему.)
4. *Степин В. С.* Саморазвивающиеся системы и постнеклассическая рациональность // Вопросы философии. 2003. № 8. С. 5–17.
5. *Вигнер Э.* Инвариантность и законы сохранения. Этюды о симметрии / Пер. Ю. А. Данилова. М.: УРСС, 2002.

6. *Пенроуз Р.* Тени разума: В поисках науки о сознании. Ч. 1. Понимание разума и новая физика / Пер. А. Р. Логунова и Н. А. Зубченко. М.; Ижевск: Ин-т компьютерных исследований, 2003.
7. *Грин Б.* Элегантная Вселенная. Суперструны, скрытые размерности и поиски окончательной теории / Пер. с англ. М.: КомКнига/URSS, 2007.
8. *Ландау Л. Д., Лифшиц Е. М.* Статистическая физика. Ч. 1. 3-е изд. М.: Наука, 1976 (первое издание вышло в 1937 г.).
9. *Чернавский Д. С.* Синергетика и информация. Динамическая теория информации. 2-е изд. М.: УРСС, 2004. (Синергетика: от прошлого к будущему.)
10. *Левкович-Маслюк Л. И.* Физическая личность // Компьютерра. 1998. № 40 (13 октября 1998 г.).
11. *Ахромеева Т. С., Курдюмов С. П., Малинецкий Г. Г., Самарский А. А.* Нестационарные структуры и диффузионный хаос. М.: Наука, 1992.
12. *Пенроуз Р.* Это — культурная революция сверху // Компьютерра. 2004. № 28 (10 августа 2004 г.).
13. *Marshall W., Simon C., Penrose R., Bouwmeester D.* Towards quantum superpositions of a mirror. http://xxx.lanl.gov/PS_cache/quant-ph/pdf/0210/0210001.pdf
14. *Куракин П. В.* Скрытые параметры и скрытое время в квантовой теории. М.: ИПМ им. М. В. Келдыша РАН, 2004. Препринт № 33.
15. *Малинецкий Г. Г., Потапов А. Б., Подлазов А. В.* Нелинейная динамика: подходы, результаты, надежды. М.: КомКнига/URSS, 2006. (Синергетика: от прошлого к будущему.)
16. *Пригожин И., Стенгерс И.* Время. Хаос. Квант. К решению парадокса времени. М.: КомКнига/URSS, 2005. (Синергетика: от прошлого к будущему.)
17. Режимы с обострением. Эволюция идеи. М.: Наука, 1999.
18. *Тьюринг А. М.* Может ли машина мыслить? / Пер. с англ. Ю. А. Данилова. М.: Гос. изд. физ.-мат. лит., 1960. 2-е изд. М.: Издательство ЛКИ/URSS (в печати).
19. *Котов Ю. Б.* Новые математические подходы к задачам медицинской диагностики. М.: УРСС, 2004. (Синергетика: от прошлого к будущему.)
20. *Владимиров В. А., Воробьев Ю. Л., Малинецкий Г. Г. и др.* Управление риском. Риск, устойчивое развитие, синергетика. М.: Наука, 2000. (Повторное изд. планируется в изд-ве УРСС в 2005 г.)
21. Новое в синергетике: Взгляд в третье тысячелетие // Информатика: неограниченные возможности и возможные ограничения / Ред. Г. Г. Малинецкий, С. П. Курдюмов. М.: Наука, 2002.
22. *Хант Э.* Искусственный интеллект. М.: Мир, 1978.
23. *Редько В. Г.* Эволюционная биокибернетика. М.: Наука, 2002.

Обращение к читателю

Как читать математические формулы

В некоторых частях этой книги я решился прибегнуть к математическим формулам. Меня не устрашило известное предостережение, что каждая формула в книге сокращает вдвое круг читателей. Если вы, Читатель, испытываете ужас перед формулами (как большинство людей), то я вам могу порекомендовать способ, который и сам часто использую, когда приличия нарушаются таким грубым образом. Способ заключается, более или менее, в том, чтобы полностью проигнорировать строку с формулой, сразу переводя взгляд на следующий за ней текст! На самом деле, конечно же, не совсем так: надо одарить формулу пытливым, но не проникающим взглядом, а затем двинуться вперед. Некоторое время спустя, почувствовав бóльшую уверенность в своих силах, можно вернуться к отвергнутой формуле и попытаться ухватить основные идеи. Текст, сопровождающий формулу, поможет вам понять, что в ней важно, а что можно спокойно проигнорировать. Если же этого все-таки не случилось, то смело оставляйте формулу и больше о ней не вспоминайте.

Посвящаю эту книгу светлой памяти моей дорогой матери, почившей прежде, чем эта книга увидела свет

Благодарности

Многие помогали мне, тем или иным способом, в написании этой книги. Всем им я очень признателен. Для начала упомяну сторонников теории сильного искусственного интеллекта (в особенности тех, которые выступали в телевизионной программе BBC), чьи радикальные идеи об искусственном интеллекте привлекли много лет назад мое внимание к этой теме. (Однако если бы я мог предвидеть заранее тот объем работы, который будет сопряжен с написанием этой книги, я вряд ли бы, думаю, начал.)

Многие скрупулезно читали отдельные части рукописи и высказывали мне свои идеи по ее улучшению. Им я приношу свою признательность. Это Тоби Бэйли, Давид Дойч (который мне очень помог в проверке описания машин Тьюринга), Стюарт Хампшир, Джим Хартли, Лэйн Хагстон, Ангус МакИнтир, Мэри Джэйн Моват, Тристан Неедман, Тед Ньюман, Эрик Пенроуз, Тоби Пенроуз, Вольфганг Риндлер, Энгельберт Шукинг и Дэннис Шьяма. Я очень благодарен Кристоферу Пенроузу за детальную информацию о множестве Мандельброта, а также Джонатану Пенроузу за сведения о шахматных компьютерах. Выражаю мою особую благодарность Колину Блэйкмору, Эрику Харту и Дэвиду Хьюбелу, которые внимательно прочитали главу 9, в предмете которой я, очевидно, совсем не специалист. Однако они — как и все остальные, которых я благодарю, — не отвечают за ошибки, если таковые сохранились. Я благодарен Национальному научному фонду за поддержку по контракту DMS 84–05644, DMS 86–06488 (университет Райса, г. Хьюстон, где проходили многие лекции, частично легшие в основу этой книги), PHY 86–12424 (университет г. Сиракузы, где я участвовал во многих ценных обсуждениях по квантовой механике). Я премного обязан Мартину Гарднеру за его великодушное предложение написать предисловие к моей книге, а также за его ценные комментарии. Особенно благодарю мою дорогую Ванессу за ее вдумчивую и детальную критику некоторых глав, за неоценимую помощь с библиографией, а также, что совсем немаловажно, за ее терпение, когда я был совсем невыносим — и за ее глубокую любовь и поддержку, когда я в этом особенно нуждался.

Предисловие

Для многих великих физиков и математиков написать книгу, понятную не только профессионалам — дело трудное, если не сказать невозможное. И вплоть до сего времени иным могло бы показаться, что Роджер Пенроуз, один из наиболее компетентных и плодотворно работающих физиков-теоретиков во всем мире, относится как раз к такой категории ученых. Но даже для тех из нас, кто был знаком с его популяризаторскими статьями и лекциями и не разделял подобного мнения, появление превосходной книги для широкого круга читателей, ради которой он оторвал от работы часть своего времени, стала приятным сюрпризом. И я не сомневаюсь, что этой книге в будущем уготовано стать классической монографией.

Хотя в различных главах своей книги Пенроуз затрагивает и теорию относительности, и квантовую механику, и космологию — главным объектом его рассуждений является так называемая психофическая проблема «ум—тело». Десятилетиями сторонники теории «сильного ИИ» (искусственного интеллекта) пытались убедить нас, что не пройдет и одного-двух веков (а некоторые опускали эту планку даже до пятидесяти лет!), как электронные компьютеры полностью сравняются по своим возможностям с человеческим мозгом. Находясь под впечатлением прочитанных в юности научно-фантастических книг и будучи убежденными в том, что наши мозги — это просто «компьютеры, сделанные из мяса» (как выразился однажды Марвин Мински), они считали несомненным, что удовольствие и боль, восприятие прекрасного и чувство юмора, сознание и свобода воли — все эти способности возникнут у электронных роботов сами собой, как только управляющие ими алгоритмы обретут достаточную степень сложности.

Но некоторые методологи науки (в особенности Джон Серл, чей мысленный эксперимент со знаменитой китайской комнатой Пенроуз очень подробно разбирает в одной из глав) с этим решительно не согласны. В их представлении компьютер по существу ничем не отличается от обычных механических калькуляторов, в которых арифметические действия выполняются посредством колесиков, рычажков или иных приспособлений, позволяющих передавать сигналы. (За основу компьютера с таким же успехом можно взять, например, маленькие перекатывающиеся шарики или текущую по системе труб воду.) Поскольку электричество движется по проводам быстрее, чем любая иная форма энергии (за исключением света), электрические устройства могут оперировать символами с большей скоростью, что позволяет им выполнять чрезвычайно громоздкие и сложные задачи. Но «осознает» ли компьютер свои действия в большей мере, чем это доступно обычным деревянным счетам? Сегодня компьютеры могут играть в шахматы на уровне гроссмейстеров. Но «понимают» ли они эту игру лучше, чем машина для «крестиков-ноликов», собранная группой компьютерных хакеров из поломанных игрушек?

Книга Пенроуза является самой мощной атакой на теорию сильного ИИ из всего написанного до сих пор. За несколько прошедших столетий было высказано немало возражений против понимания мозга как машины, управляемой общеизвестными

законами физики; но доводы Пенроуза более убедительны, ибо они базируются на недоступной для его предшественников информации. Эта книга открывает нам другого Пенроуза — не только математика и физика, но и философа высокого уровня, не отступающего перед проблемами, которые современные философы слишком легко сбрасывают со счетов как бессмысленные.

К тому же Пенроуз, вопреки все более настойчивым возражениям небольшой группы физиков, имеет смелость отстаивать позиции здорового реализма. В его представлении реальна не только вселенная, но и математическая истина, непостижимым образом ведущая свое собственное независимое и вечное существование. Подобно Ньютону и Эйнштейну, Пенроуз испытывает благоговейный трепет и чувство смирения как перед физическим миром, так и перед Платоновым царством чистой математики. Выдающийся ученый в области теории чисел Пол Эрдос любит говорить «о божественной книге», в которой записаны все лучшие доказательства. И математикам иной раз приоткрывается та или иная ее страница. Моменты прозрения, когда математик или физик внезапно вскрикивает «Ага!», по мнению Пенроуза, не могут явится «результатом сколь угодно сложных вычислений»: в эти мгновения разум соприкасается с объективной истиной. Возможно ли, вопрошает Пенроуз, что мир «идей» Платона и реальный физический мир (который физики сегодня все больше «растворяют» в математике) — на самом деле тождественны?

Большое внимание в книге Пенроуза уделяется знаменитой фрактальной структуре, называемой множеством Мандельброта в честь ее первооткрывателя Бенуа Мандельброта. Хотя в статистическом смысле такие объекты обладают свойством самоподобия, которое выявляется при увеличении отдельных частей, их бесконечно причудливые очертания постоянно меняются самым непредсказуемым образом. Пенроузу кажется непонятным, как можно сомневаться в том, что эти экзотические структуры существуют не менее «реально», чем гора Эверест, и могут быть исследованы точно так же, как исследуются джунгли.

Пенроуз принадлежит к постоянно пополняющейся группе ученых, которые считают, что Эйнштейн не был упрямым или, тем более, бестолковым, когда однажды, ссылаясь на свой «левый мизинец», он провозгласил неполноту квантовой механики. Чтобы подтвердить справедливость этого утверждения, Пенроуз увлекает читателя в головокружительное путешествие, в ходе которого мы знакомимся с комплексными числами, машинами Тьюринга, теорией сложности, поразительными парадоксами квантовой механики, формальными системами, (теоремой) неразрешимости Геделя, фазовыми и гильбертовыми пространствами, черными и белыми дырами, излучением Хокинга, энтропией, строением мозга — и множеством других вопросов, занимающих сегодня умы ученых. «Осознают» ли кошки и собаки свое «я»? Могут ли в теории существовать передатчики материи, способные переместить человека из одного места в другое на манер астронавтов из сериала *Звездный Путь*? Насколько полезно нам — с точки зрения выживания — возникшее в ходе эволюции сознание? Существует ли структура более общая, чем квантовая механика, где бы нашлось естественное объяснение направлению времени и различиям между правым и левым? Важны ли законы квантовой механики, а может и некие более «тонкие» законы, для деятельности разума?

На два последних вопроса Пенроуз дает положительный ответ. Его знаменитая теория «твисторов» — абстрактных геометрических объектов, действующих в многомерном комплексном пространстве, которое лежит в основе обычного пространства-времени — носит чересчур узкоспециализированный характер, чтобы быть включенной в эту книгу. Она стала результатом его двадцатилетних усилий проникнуть в область более глубокую, чем квантовые поля и частицы. Прибегая к своей четырехступенчатой классификации теорий — превосходных, полезных, пробных и тупиковых, — Пенроуз скромно поместил теорию твисторов в разряд пробных, вместе с суперструнами и другими теориями великого объединения, которые сейчас вызывают острые дискуссии в научной среде.

С 1973 года Пенроуз возглавляет кафедру Рауза Болла в Оксфордском университете. Это тем более заслуженно, что В. У. Рауз Болл был не только выдающимся математиком, но еще и фокусником-любителем, настолько увлеченным занимательной математикой, что однажды он даже написал на эту тему ставшую классической книгу *Математические эссе и развлечения*[1]. Пенроуз разделяет эту страсть Болла к играм. В юности он придумал «невозможный объект», состоящий из трех стержней. (Невозможный объект — это изображение цельной фигуры, которая не может существовать из-за наличия в ней внутренне противоречивых элементов.) Вместе со своим отцом Лайонелом, генетиком по профессии, он превратил свой невозможный объект в *Лестницу Пенроуза*, структуру, использованную Морицем Эшером на двух известных литографиях: *Идущие вверх и идущие вниз* и *Водопад*. В один прекрасный день, когда Пенроуз лежал в кровати, с ним случился, как он сам называет это, «приступ сумасшествия», когда ему явственно представился невозможный объект в четырехмерном пространстве. Если бы существо из четырехмерного мира наткнулось на эту штуку, шутит Пенроуз, оно наверняка воскликнуло бы: «Боже мой, что это такое!?»

Работая в 1960-х годах вместе со своим другом Стивеном Хокингом над проблемами космологии, он сделал свое самое, наверное, известное открытие. Если теория относительности выполняется «до самого конца», то в каждой черной дыре должна существовать сингулярность, где законы физики теряют свою силу. Но даже это достижение отошло в последние годы на второй план, после того как Пенроуз предложил конструкцию из «плиток» двух видов, которыми можно покрыть всю плоскость подобно мозаике Эшера — только непериодическим образом. (Об этих удивительных фигурах вы можете узнать подробнее в моей книге *От мозаик Пенроуза к надежным шрифтам*[2].) Пенроуз изобрел, или, скорее, открыл их, даже не предполагая, что когда-нибудь они могут кому-то пригодиться. К всеобщему изумлению оказалось, что трехмерные аналоги этих фигур могут служить основой для новой необычной формы материи — «квазикристаллов». Сейчас изучение «квазикристаллов» превратилось в одну из наиболее активных областей исследований в кристаллографии. Это, безусловно, самый впечатляющий пример того, как в наши дни математические игры могут иметь совершенно неожиданные практические приложения.

Достижения Пенроуза в математике и физике — а я упомянул только незначительную их часть — рождаются из постоянно присутствующего в его душе ощущения тайны и красоты бытия. Мизинец «подсказывает» ему, что человеческий мозг представляет собой устройство более сложное, чем набор крошечных проводков и переключателей. Фигура Адама в прологе и эпилоге этой книги в определенном смысле служит символом зарождения разума в ходе неторопливого развития осознающей себя жизни. В нем я тоже вижу Пенроуза — мальчика, сидящего в третьем ряду, позади признанных корифеев в области ИИ, — который не боится высказать им вслух свое мнение, что их «короли-то голые»[3]. Юмор присущ многим высказываниям Пенроуза, но это утверждение — отнюдь не шутка.

Мартин Гарднер

---

[1] Rouse Ball, W. W. and Coxeter, H. S. M., *Mathematical Recreations and Essays*, Macmillan, London, 1959. (Рус. пер.: *Болл У., Коксетер Г.* Математические эссе и развлечения. М.: Мир, 1986). — *Прим. ред.*

[2] Gardner, M., *Penrose tiles to trapdoor ciphers*. W. H. Freeman and Company, New York, 1989. (Рус. пер.: *Гарднер М.* От мозаик Пенроуза к надежным шрифтам. М.: Мир, 1993). — *Прим. ред.*

[3] В оригинале название книги *The Emperor's New Mind* перекликается с названием известной сказки Г.-Х. Андерсена *The Emperor's New Clothes* — *Новый наряд короля*. — *Прим. ред.*

Вступление

Книга *Новый ум короля*, впервые изданная в 1989 году, стала моей первой серьезной попыткой написать научно-популярное произведение. Приступая к созданию этой книги, я, помимо всего прочего, ставил целью рассказать в максимально доступной форме о значительном прогрессе физической науки, достигнутом в познании законов окружающего нас мира. Но это не просто обзор научных достижений. Я еще и пытаюсь указать на целый ряд принципиальных трудностей, которые стоят перед наукой на ее пути к конечной цели. В частности, я утверждаю, что явление *сознания* не может быть описано в рамках современной физической теории.

Это явно противоречит довольно устоявшемуся пониманию сущности научного подхода, согласно которому *все* аспекты умственной деятельности (включая, в том числе, и сознание) — не более, чем результат вычислений, происходящих в мозге; соответственно, электронные компьютеры должны быть потенциально способны к сознательному восприятию, которое возникло бы само собой при наличии достаточной мощности и соответствующих программ. Я постарался по возможности беспристрастно аргументировать свое несогласие с таким взглядом, указывая на то, что проявления сознательной деятельности мозга не могут быть объяснены в вычислительных терминах и — более того — с позиций современного научного мировоззрения в целом. Однако я ни в коем случае не утверждаю, что понимание этого феномена невозможно в рамках научного подхода — просто современная наука еще не достигла уровня, необходимого для решения такой задачи.

Когда я писал эту книгу, мне трудно было вообразить, сколь бурной окажется реакция на изложенные в ней мысли — причем не только из лагеря убежденных сторонников «компьютерной» модели разума, но и со стороны тех, кто считает научный метод недопустимым для изучения сознания. Я нисколько не сомневаюсь, что попытка затронуть чью-то личную философскую концепцию сознания — как и религиозные воззрения — может оказаться делом довольно рискованным. Но *насколько* щекотливой бывает подчас эта тема — я едва ли мог представить себе в полной мере.

Мои рассуждения в том виде, в котором они представлены в книге, направлены на достижение двух целей. Первая из них — это стремление показать, опираясь главным образом на результаты, полученные Геделем (и Тьюрингом), что математическое мышление — а, следовательно, и умственная деятельность в целом — не может быть полностью описано при помощи чисто «компьютерной» модели разума. Именно эта часть моих умозаключений вызывает у критиков наиболее настойчивые возражения. Вторая цель — показать, что сегодня в физической картине мира есть существенное «белое пятно», а именно: отсутствует «мостик» между субмикроскопическим уровнем квантовой механики и макромиром классической физики. С моей точки зрения, теория, которая однажды восполнит этот пробел, должна будет в значительной степени помочь понять физические основы феномена сознания. Более того, в этой искомой области физики должно быть заложено нечто выходящее за рамки только вычислительных действий.

За десятилетие, прошедшее с момента первого издания книги, наука добилась целого ряда ошеломляющих успехов. Про некоторые из них я бы хотел вкратце рассказать здесь с тем, чтобы у читателя сложилось определенное представление о моем вйдении современного состояния этих исследований. Сперва рассмотрим, насколько важна теорема Геделя для критики выдвинутых мной положений. Если попытаться изложить в двух словах суть этой теоремы (справедливость которой не оспаривается), то она будет выглядеть следующим образом. Пусть мы располагаем какой-нибудь вычислительной процедурой $\boldsymbol{P}$, позволяющей нам формулировать математические утверждения (для определенности договоримся, что это будут утверждения какого-то одного вида, аналогичные, допустим, знаменитой теореме Ферма (см. с. 82)). Тогда, если мы готовы считать правила процедуры $\boldsymbol{P}$ *надежными* — в том смысле, что мы будем полагать всякое математическое утверждение, полученное при помощи этой процедуры, *неоспоримо* верным, — то равным образом мы должны принимать и неоспоримую справедливость некоторого утверждения $G(\boldsymbol{P})$, которое *лежит за пределами действия* правил процедуры $\boldsymbol{P}$ (см. с. 115). Таким образом, как только мы научились автоматизировать некоторую часть нашего математического мышления, у нас сразу же появляется понимание, как выйти за его границы. В моем представлении это однозначно свидетельствует о том, что математическое понимание содержит определенные элементы, которые не могут быть полностью сведены к вычислительным методам. Но многие критики остались при своих убеждениях, указывая на различные возможные «тонкие места» в этих логических построениях. В моей следующей книге *Тени разума*[1] я постарался ответить на все подобные возражения и привел ряд новых аргументов в пользу своей точки зрения. Тем не менее споры все еще продолжаются[2].

Одна из причин, мешающих людям признать прямое отношение, которое имеет теорема Геделя к нашему математическому мышлению, заключается в том, что в рамках обычной ее формулировки утверждение $G(\boldsymbol{P})$ не представляет интереса с математической точки зрения. Мало того: оно еще и чрезвычайно сложно для понимания в качестве математического выражения. Соответственно, даже математики предпочитают не «связываться» с подобными выражениями. Однако, существует ряд примеров утверждений геделевского типа, которые легко доступны пониманию даже для тех, чье знакомство с математической терминологией и системой записи ограничивается рамками обычной арифметики.

Особенно впечатляющий пример попался мне на глаза уже после того, как была опубликована эта книга (а также *Тени разума*). Это произошло на лекции Дэна Исаксона в 1996 году. Речь шла об известной *теореме Гудстейна*[3]. Данный пример кажется мне настолько поучительным, что я хотел бы рассмотреть его здесь целиком, дабы читатель имел возможность непосредственно познакомиться с теоремами геделевского типа[4].

Чтобы понять суть этой теоремы, рассмотрим любое целое положительное число, скажем, 581. Для начала мы представим его в виде суммы различных степеней числа 2:

$$581 = 2^9 + 2^6 + 2^2 + 1.$$

(Такая процедура применяется для формирования двоичного представления числа 581, а именно, приведения его к виду 1001000101, где единицы соответствуют тем степеням двойки, которые присутствуют в таком представлении, а нули — тем степеням, которых

---

[1] *Shadows of the Mind*, Oxford University Press, 1994; pb. Vintage, 1995.

[2] Заинтересованный читатель может ознакомиться с критическими замечаниями и моими ответами на них в *Behavioral and Brain Sciences*, 13 (4) (1990), 643–705 и в *Psyche* (MIT Press), 2 (1996), 1–129. Последний из этих материалов можно найти на веб-сайте http://psyche.cs.monash.edu.au/psyche-index-v2_1.html; я рекомендую прочитать приведенные мной возражения (озаглавленные *Beyond the Doubting of a Shadow*) на критические замечания по этому вопросу перед тем, как приступить к чтению книги *Тени разума*. Следующим источником дополнительной информации может служить моя работа *The Large, the Small and the Human Mind* (Cambridge University Press, 1997).

[3] Goodstein, R. L., On the restricted ordinal theorem. *Journal of Symbolic Logic*, 9 (1944), 33–41.

[4] См. также Penrose, R., On understanding understanding. *International Studies in the Philosophy of Science*, 11 (1997), 7–20.

нет.) Далее можно заметить, что «показатели» в этом выражении — т. е. 9, 6 и 2 — могут быть, в свою очередь, представлены аналогичным образом ($9 = 2^3 + 1$, $6 = 2^2 + 2^1$, $2 = 2^1$); и тогда мы получим (вспоминая, что $2^1 = 2$)

$$581 = 2^{2^3+1} + 2^{2^2+2} + 2^2 + 1.$$

Здесь все еще есть показатель больший, чем двойка — в данном случае это «3», — для которого тоже можно написать разложение $3 = 2^1 + 1$, так что в конце концов мы будем иметь

$$581 = 2^{2^{2^1+1}+1} + 2^{2^2+2} + 2^2 + 1.$$

А теперь мы подвергнем это выражение последовательности чередующихся простых операций, которые будут

(а) увеличивать «основание» на единицу,

(б) вычитать единицу.

Под «основанием» здесь понимается просто число «2», фигурирующее в исходном выражении, но мы можем сделать то же самое и с бóльшими основаниями: 3, 4, 5, 6, ... Давайте посмотрим, что произойдет при применении операции (а) к последнему разложению числа 581, в результате которой двойки становятся тройками:

$$3^{3^{3^1+1}+1} + 3^{3^3+3} + 3^3 + 1$$

(что дает — если выписать его в обычной форме — сороказначное число, начинающееся с 133027946...). После этого мы применяем (б) и получаем

$$3^{3^{3^1+1}+1} + 3^{3^3+3} + 3^3$$

(т. е. по-прежнему сорокозначное число, начинающееся с 133027946...). Далее мы выполняем (а) еще раз и получаем

$$4^{4^{4^1+1}+1} + 4^{4^4+4} + 4^4$$

(это уже значительно большее число, состоящее из 618 знаков, которое начинается с 12926802...). Следующая операция — вычитание единицы — приводит к выражению

$$4^{4^{4^1+1}+1} + 4^{4^4+4} + 3 \times 4^3 + 3 \times 4^2 + 3 \times 4 + 3$$

(где тройки получаются по той же причине, что и девятки в обычной десятичной записи, когда мы получаем 9999, вычитая 1 из 10 000). После чего операция (а) дает нам

$$5^{5^{5^1+1}+1} + 5^{5^5+5} + 3 \times 5^3 + 3 \times 5^2 + 3 \times 5 + 3$$

(число, которое имеет 10 923 знака и начинается с 1274...). Обратите внимание, что коэффициенты «3», которые возникают при этом, с необходимостью меньше, чем основание (в данном случае 5), и не изменяются с возрастанием последнего. Применяя (б) вновь, имеем число

$$5^{5^{5^1+1}+1} + 5^{5^5+5} + 3 \times 5^3 + 3 \times 5^2 + 3 \times 5 + 2,$$

над которым мы опять производим последовательно действия (а), (б), (а), (б), ... и т. д., насколько возможно. Вполне естественно предположить, что этот процесс никогда не завершится, потому что каждый раз мы будем получать все бóльшие и бóльшие числа. Однако это не так: как следует из поразительной теоремы Гудстейна, независимо от величины исходного числа (581 в нашем примере), мы *в конце концов получим нуль!*

Кажется невероятным, но это так. А чтобы в это поверить, я рекомендовал бы читателю самостоятельно проделать вышеописанную процедуру, для начала — с числом «3» (где мы раскладываем тройку как $2^1 + 1$, что дает последовательность 4, 3, 4, 2, 1, 0); а затем — что более важно — попробовать то же самое с «4» (при этом стартовое разложение в виде $4 = 2^2$ приводит к вполне закономерно возрастающему ряду 4, 27, 26, 42, 41, 61, 60, 84, ..., который доходит до числа из 121 210 695-ти знаков, после чего уменьшается вплоть до нуля!).

Но что кажется еще более удивительным: теорема Гудстейна фактически является *теоремой Геделя* для той самой процедуры, которую мы изучали в школе под названием *математической индукции*, как было доказано в свое время Л. Кирби и Дж. Парисом[5]. Как вы, должно быть, помните, математическая индукция позволяет установить справедливость некоторого математического утверждения $S(n)$ для $n = 1, 2, 3, 4, 5, \ldots$. Доказательство проводится в два этапа: сначала нужно проверить справедливость S(1),

[5] Accessible independence results for Peano arithmetic. *Bulletin of the London Mathematical Society*, **14** (1982), 285–93.

а затем показать, что, если верно $S(n)$, то должно выполняться и $S(n+1)$. Приняв процедуру математической индукции за $\boldsymbol{P}$, Кирби и Парис доказали, что тогда $G(\boldsymbol{P})$ может иметь смысл теоремы Гудстейна.

Следовательно, если мы считаем процедуру математической индукции достоверной (с чем едва ли можно не согласиться), то мы должны верить и в справедливость теоремы Гудстейна — несмотря на то, что при помощи одной лишь математической индукции доказать ее *невозможно.*

«Недоказуемость» теоремы Гудстейна, понимаемая в этом смысле, вряд ли может помешать нам убедиться в ее фактической *справедливости.* Наши интуитивные представления позволяют нам *расширить* действие тех ограниченных приемов «доказательства», которыми мы воспользовались ранее. В действительности сам Гудстейн доказал свою теорему, прибегнув к разновидности метода, который называется «трансфинитной индукцией». В контексте нашего изложения этот метод сводится к систематизации интуитивных ощущений, которые возникают в процессе знакомства с «причиной», по которой теорема Гудстейна и в самом деле верна. Эти ощущения могут родиться практически целиком за счет изучения некоторого числа частных случаев указанной теоремы. И тогда станет видно, как скромная незаметная операция (б) безжалостно «отщипывает» по кусочку от огромной башни «показателей» до тех пор, пока она не начинает постепенно таять и полностью исчезает, — хотя бы на это ушло и невообразимо большое число шагов.

Все это говорит о том, что способность *понимать* никоим образом не может сводиться к некоторому набору правил. Более того, понимание является свойством, которое зависит от нашего сознания; и что бы не отвечало в нас за сознательное восприятие — это должно самым непосредственным образом участвовать в процессе «понимания». Тем самым, в формировании нашего сознания с необходимостью есть элементы, которые не могут быть получены из какого бы то ни было набора вычислительных инструкций; что, естественно, дает нам веские основания считать, что сознательное восприятие — процесс существенно «невычислимый».

Возможные «узкие места» в этом рассуждении сводятся к следующему. Наша способность (математического) познания может быть результатом вычислительной процедуры или непознаваемой из-за своей сложности; или не непознаваемой, но правильность которой, однако, не может быть установлена; или же ошибочной, хотя почти правильной. Говоря об этом, мы должны прежде всего установить, откуда могут возникать подобные вычислимые процедуры. В книге *Тени разума* я достаточно подробно рассмотрел все такие «узкие места», и я хотел бы порекомендовать эту книгу (равно как и статью *Beyond the Doubting of a Shadow* в журнале *Psyche*[6]) всем читателям, кому интересно было бы ближе познакомиться с настоящим предметом.

Если мы согласимся с тем, что в нашей способности познавать — а следовательно, и в нашей сознательной деятельности в целом — есть нечто, выходящее за пределы чисто алгоритмических действий, то следующим шагом мы должны попытаться выяснить, в каких из наших физических действий может проявляться «существенно неалгоритмическое поведение». (При этом мы негласно предполагаем, что изучение именно «физического действия» определенного вида поможет нам разгадать тайну происхождения сознания.) Я пытаюсь доказать, что таким «неалгоритмическим действиям» нельзя найти место в рамках общепринятых сегодня физических теорий. А значит, мы должны искать соответствующее место, где в научной картине существует серьезный *пробел.* И я утверждаю, что это «белое пятно» лежит где-то на границе между «субмикроскопическим» миром, в котором правит квантовая механика, и непосредственно воспринимаемым нами макромиром, подчиняющимся законам классической физики.

Здесь необходимо сделать важное замечание. Термин «невычислимый» относится к некоторому классу математических действий, про которые известно — то есть *доказано* математически, — что они не поддаются вычислениям. И одна из задач данной

---

[6] См. сноску 2 на с. 32.

книги заключается в том, чтобы познакомить читателя с этим вопросом. Невычислимые процессы могут быть полностью детерминистскими. Эта особенность является диаметрально противоположной по отношению к свойству полной *случайности*, которое характерно для современной интерпретации квантовой механики и возникает при увеличении микромасштабных квантовых эффектов до классического уровня — R-процедуре в моей терминологии в этой книге. Я считаю, что необходима новая теория, которая позволит постичь смысл «реальности», принадлежащей сфере действия R-процедуры, которая сегодня используется в квантовой механике; и, как мне кажется, именно в этой неоткрытой пока новой теории мы найдем требуемый элемент невычислимости.

Кроме того, я смею утверждать, что эта недостающая теория является одновременно и искомым звеном между квантовой механикой и общей теорией относительности Эйнштейна. Для этой единой теории в физике применяется название «квантовая гравитация». Однако, большинство работающих в этой области ученых полагают, что объединение двух величайших теорий двадцатого века не затронет законов квантовой механики, в то время как общая теория относительности должна претерпеть изменения. Я придерживаюсь иной точки зрения, поскольку считаю, что методы квантовой теории (в частности, R-*процедура*) тоже должны существенно измениться. В этой книге я использовал термин «правильная квантовая теория гравитации» (или «ПКТГ»), чтобы обозначить возможный результат такого объединения — хотя это и не будет теорией квантовой гравитации в обычном смысле (и, вероятно, «ПКТГ» тоже не очень удачный термин, который может ввести кого-то в заблуждение).

Хотя такой теории до сих пор не существует, это вряд ли может помешать нам оценить уровень, на котором она становится применимой. В книге я использовал для этих целей «одногравитонный критерий». Но несколько лет спустя я был вынужден изменить свои взгляды и, как мне кажется, найти более адекватный подход, изложенный в книге *Тени разума*. Этот подход близок к реальности не только «физически» (чему нашлось дополнительное подтверждение, которое я привел в одной из своих статей)[7], но и с практической точки зрения, что подтолкнуло нас к дальнейшим теоретическим изысканиям. На самом деле, сейчас уже разработан ряд физических экспериментов, которые, надеюсь, можно будет осуществить в ближайшие несколько лет[8].

Но даже если все перечисленное окажется справедливым и мои умозаключения подтвердятся, это не поможет нам отыскать «местоположение сознания». Вероятно, один из недостатков этой книги заключается в том, что к моменту завершения работы над ней я так и не знал, в каком месте мозга может происходить «крупномасштабная квантовая когерентность», которая необходима для использования приведенных выше идей. С другой стороны, к достоинствам книги следует отнести то, что она вызвала живой интерес в самых широких научных кругах, представители которых могут внести ценный вклад в исследования этого вопроса. Одним из таких ученых оказался Стюарт Хамерофф, который познакомил меня с цитоскелетом клетки и входящими в него микротрубочками — структурами, о которых я, к сожалению, не имел ни малейшего представления! Он также изложил мне свои оригинальные идеи по поводу возможной роли микротрубочек в нейронах мозга для феномена *сознания* — что позволило мне предположить, что они-то и являются скорее всего тем местом, где может происходить крупномасштабная квантовая когерентность, на которую я опирался в своих рассуждениях. Конечно же, эта информация достигла меня уже слишком поздно, чтобы я мог включить ее в *настоящее* издание; но ее изложение можно найти в книге *Тени разума* и последующих статьях, написанных преимущественно в соавторстве со Стюартом Хамероффом[9].

---

[7] On the gravity role in quantum state reduction, *General Relativity and Gravitation*, 28 (1996), 581–600.

[8] См. Penrose, R., Quantum computation, enanglement and state reduction, *Phil. Trans. Royal Soc. London*, A356 (1998), 1927–39; и Moroz, I., Penrose, R. and Tod, K. P., Spherically symmetric solutions of the Schrödinger—Newton equations, *Classical and Quantum Gravity*, 15 (1998).

[9] Hameroff, S. R. and Penrose, R., Conscious events as orchestrated space-time selections, *J. Consciousness Studies*,

Кроме последних достижений, упомянутых в этом новом вступлении, можно сказать, что все основные идеи книги *Новый ум короля* сохранились в том же виде, что и десять лет назад. Я надеюсь, что читатель, познакомившись с изложенными здесь мыслями, получит неподдельное удовольствие и почувствует желание самостоятельно продолжить изучение этих вопросов.

Роджер Пенроуз
Сентябрь 1998

---

3 (1996), 36–63. Hameroff, S. R. and Penrose, R., Orchestrated reduction of quantum coherence in brain microtubules — a model for consciousness; в сборнике *Towards a science of consciousness: contributions from the* 1994 *Tucson Conference* (ed. S. Hameroff, A. Kazniak and A. Scott), MIT Press 1996. Hameroff, S. R., Fundamental Geometry: the Penrouse—Hameroff "Orch OR" model of cosciousness в сборнике *The geometric universe: science, geometry and the work of Roger Penrose* (ed. S. A. Huggett, L. J. Mason, K. P. Tod, S. T. Tsou and N. M. J. Woodhouse), Oxford University Press, 1998.

Пролог

На церемонию запуска нового компьютера *Ультроник* в Большой аудитории собралась огромная толпа. Президент Полло только что закончил свое вступительное слово. Он рад, что наконец отделался — подобные мероприятия ему не по вкусу, а в компьютерах ему интересно лишь одно: эта новая штуковина позволит ему сэкономить кучу времени. Разработчики уверяли его, что помимо всего прочего, *Ультроник* будет способен отвечать за принятие решений в государственных делах, которые всегда докучали президенту. И неплохо бы, чтобы это оказалось правдой — учитывая то, сколько за этот компьютер заплачено золота из казны! Президент уже предвкушал многочасовые игры в гольф на своем личном поле — одном из немногих оставшихся в его крохотной стране островков зелени.

Адаму лестно находиться среди приглашенных на церемонию открытия. Он сидит в третьем ряду; через два ряда впереди него сидит его мать, главный технолог разработки *Ультроник*. Вышло так, что и отец его тоже находится здесь: он пришел без приглашения и сидит сейчас в самом конце зала, окруженный со всех сторон охранниками — и все потому, что в последний момент отец решил взорвать компьютер. Он сам поручил себе это задание как доморощенный лидер маленькой группы маргинальных активистов, именующей себя Высший Совет Психического Самосознания. Конечно, всю его взрывчатку тут же обнаружили установленные в изобилии электронные и химические датчики, и в качестве наиболее приятной части предстоящего наказания ему довелось стать невольным свидетелем церемонии запуска.

Адам не испытывал особых чувств ни к одному из своих родителей. Быть может, в таких чувствах у него и не было необходимости: все тринадцать лет своей жизни он рос в атмосфере материальной роскоши, обусловленной в основном возможностями компьютеров. Любое свое желание он мог удовлетворить простым нажатием на кнопку мыши — будь то потребность в еде, питье, компании или развлечениях, а если нужно, то и в знаниях — и всегда это сопровождалось прекрасными цветными иллюстрациями на графических мониторах. Все было возможно благодаря положению, которое занимала мать Адама.

И вот Главный конструктор проекта уже заканчивает свой доклад: «...более 10^{17} логических ячеек. Это больше, чем суммарное число нейронов у всех живущих в нашей стране! Уровень интеллекта невообразимо высок. Но, к счастью, нам и не нужно ничего воображать — через минуту у каждого будет возможность убедиться в этом собственными глазами! Я попрошу уважаемую первую леди нашей великой страны, мадам Изабеллу Полло, включить рубильник питания нашего фантастического компьютера *Ультроник*!»

Супруга президента подается вперед. Немного нервничая и чуть колеблясь, она поворачивает рубильник. Небольшой шорох, еле ощутимое мерцание индикаторов — и вот, 10^{17} логических ячеек активированы! Все замерли в ожидании, не совсем представляя, чего собственно и ожидать. «Итак, найдется в этой аудитории желающий инициировать нашу новую компьютерную систему *Ультроник*, задав ей первый вопрос?» — обращается к залу Главный конструктор.

Всеобщая растерянность. Никто не решается, дабы не оказаться глупцом при таком скоплении народа — и перед новым Вездесущим Разумом. Тишина. «Ну что же вы, наверняка кто-то хочет задать вопрос!» — не сдается Главный конструктор. Все в смятении, как будто чувствуя присутствие нового всемогущего разума. Лишь Адам хладнокровен. Он окружен компьютерами с самого рождения. Он почти чувствует, что значит *быть* компьютером. Или, по крайней мере, ему так кажется. Во всяком случае, он заинтригован. Адам поднимает руку. «Ну вот, — говорит Главный конструктор, — парнишка в третьем ряду. У тебя есть вопрос к нашему... гм... нашему новому другу?»

Глава 1

Может ли компьютер обладать разумом?

Введение

На протяжении нескольких предыдущих десятилетий компьютерные технологии развивались семимильными шагами. Более того, нет никаких сомнений в том, что и будущее сулит нам новые грандиозные успехи в повышении быстродействия и объема памяти, а также новые конструктивные решения компьютерной логики. Сегодняшние компьютеры завтра покажутся нам такими же медленными и примитивными, как механические калькуляторы прошлого. В таком стремительном развитии есть что-то почти пугающее. Уже сейчас машины способны решать различные задачи, ранее являвшиеся исключительной прерогативой человеческого интеллекта. И решать их со скоростью и точностью, во много раз превосходящими человеческие способности. Мы давно свыклись с существованием устройств, превосходящих наши *физические* возможности. И *это* не вызывает у нас внутреннего дискомфорта. Наоборот, нам более чем комфортно, когда автомобиль несет нас в пять раз быстрее, чем лучший в мире бегун. Или когда с помощью таких устройств мы копаем ямы или сносим непригодные конструкции — с эффективностью, которую не разовьет и отряд из нескольких дюжин добрых молодцев. Еще больше нам импонируют машины, с помощью которых у нас появляется возможность делать то, что нам ранее было попросту недоступно физически, например, подняться в небо и всего через несколько часов приземлиться на другом берегу океана. Эти машины не задевают нашего тщеславия. Но вот способность *мыслить* всегда была прерогативой человека. В конце концов, именно этой способности мы обязаны тому, что человеку удалось преодолеть его физические ограничения и встать в развитии на ступеньку выше над другими живыми существами. А если когда-нибудь машины превзойдут нас там, где, по нашему мнению, нам нет равных — не получится ли так, что мы отдадим пальму первенства своим же собственным творениям?

Можно ли считать, что механическое устройство в принципе способно мыслить, или даже испытывать определенные чувства? Этот вопрос не нов [1], но с появлением современных компьютерных технологий он приобрел новое значение. Смысл вопроса глубоко философский. Что значит — думать или чувствовать? Что есть разум? Существует ли он объективно? И если да, то в какой степени он функционально зависим от физических структур, с которыми его ассоциируют? Может ли он существовать независимо от этих структур? Или он есть лишь продукт деятельности физической структуры определенного вида? В любом случае — должны ли подходящие структуры быть обязательно биологическими (мозг) или, возможно, этими структурами могут быть и электронные устройства? Подчиняется ли разум законам физики? И вообще, *что* такое законы физики?

Вот часть проблем, которые я попытаюсь затронуть в этой книге. Просить дать определенный ответ на такие глобальные

вопросы — это, конечно, было бы слишком. Я не способен дать такой ответ, да и никто не способен — хотя некоторые, возможно, попытались бы вас обескуражить своими догадками. Мои собственные догадки играют большую роль в последующем изложении, но я постараюсь очень внимательно подчеркивать, где кончается строгий научный анализ и начинаются догадки, а также то, чем мои соображения мотивированы. Я не пытаюсь угадать правильные ответы: моя главная задача куда скромнее. Цель этой книги — поднять ряд, по-видимому, новых вопросов о взаимосвязи структуры физических законов, естества математики и разумного мышления, а также представить точку зрения, отличную от тех, которые я когда-либо встречал. Я не могу описать эту точку зрения в двух словах — вот одно из объяснений того, почему я решил написать книгу такого объема. Но если суммировать кратко (хотя краткость вполне может ввести читателя в заблуждение), моя позиция основана на осознании того, что именно наше недостаточное понимание фундаментальных физических законов препятствует построению концепции «разума» в физических и логических терминах. Я не утверждаю, что мы никогда не познаем физические законы в достаточной для этого степени. Наоборот, одна из задач книги — попытаться дать стимул дальнейшим исследованиям в наиболее перспективных в данном отношении направлениях, и попробовать пояснить достаточно определенные (и, вероятно, свежие) соображения о месте, которое могло бы занимать понятие «разума» в известной нам физической науке.

Сразу отмечу, что моя точка зрения не является общепринятой среди физиков. Поэтому маловероятно, что в настоящее время она получит признание ученых-компьютерщиков или психологов. Любой физик скажет вам, что фундаментальные законы, действующие на масштабах, характерных для человеческого мозга, прекрасно известны. Хотя никто не отрицает, что в наших знаниях физики как таковой многого недостает. Мы, например, не знаем ни основных законов, которые определяют значения масс субатомных частиц, ни законов, определяющих силу взаимодействия между этими частицами. Мы не знаем, как добиться полного согласования квантовой теории и специальной теории относительности Эйнштейна — не говоря уже о том, как построить теорию квантовой гравитации, в рамках которой удалось бы согласовать квантовую теорию и *общую* теорию относительности. Вследствие этого мы не способны понять природу пространства на чрезвычайно малых расстояниях порядка 1/100 000 000 000 000 000 000 размеров известных фундаментальных частиц, хотя и считается, что на бóльших расстояниях наши представления являются адекватными. Мы не знаем, является ли вселенная как единое целое конечной или бесконечной в пространственных или во временнóм измерениях, хотя подобные неопределенности, по-видимому, совершенно несущественны для физики важных для человека явлений. Мы не представляем себе, какие физические законы работают в сердцевине черных дыр и какие законы действовали в момент Большого взрыва при рождении самой нашей вселенной. Все перечисленные проблемы, однако, кажутся нам невообразимо далекими от шкалы явлений «повседневной» жизни (или чуть меньшей шкалы), от масштабов, характерных для жизнедеятельности человеческого мозга. И эти проблемы действительно невообразимо далеки! Тем не менее, я утверждаю, что в нашем понимании физического мира есть брешь *именно на том* уровне, который может иметь непосредственное отношение к работе человеческого мозга и сознанию. Эта брешь — прямо у нас под носом (или, скорее, за ним)! Однако большинство физиков даже не чувствуют ее — ниже я попытаюсь объяснить почему. Далее я приведу доводы в пользу того, что теории черных дыр и Большого взрыва на самом деле *имеют* определенное отношение к рассматриваемым вопросам!

Ниже я постараюсь убедить читателя в силе рассуждений, лежащих в основе предлагаемой мною точки зрения. Но чтобы понять ее, потребуется изрядно потрудиться. Нам понадобится совершить путешествие в довольно странные области (кажущиеся, возможно, не имеющими отношения к делу) и заглянуть во многие сферы научной деятельности. Будет необхо-

димо подробно изучить структуру, основы и парадоксы квантовой теории, основные положения специальной и общей теории относительности, теории черных дыр, Большого взрыва, второго закона термодинамики, максвелловской теории электромагнитных явлений, а также основы механики Ньютона. При попытке понять природу и работу сознания в игру немедленно войдут также философия и психология. Имея перед собой компьютерные модели, мы, конечно, не обойдемся и без экскурса в нейрофизиологию живого мозга. Нам понадобится также некоторое представление о статусе искусственного интеллекта. Потребуется разобраться, что такое машина Тьюринга, понять смысл вычислимости, теоремы Геделя и теории сложности. Кроме того, нам придется окунуться в дебри оснований математики и даже обсудить вопрос о самой природе физической реальности.

И если после всего этого читатель останется скептически настроен к наиболее необычным из моих аргументов, то мне, по крайней мере, хочется верить, что он вынесет нечто действительно ценное из этого изматывающего, но (я надеюсь) увлекательного путешествия.

Тест Тьюринга

Представьте себе, что появилась новая модель компьютера, объем памяти и число логических ячеек которого больше, чем у человеческого мозга. Представьте далее, что такие компьютеры грамотно запрограммированы и в них введено огромное количество необходимых данных. Производители убеждают вас, что эти устройства могут на самом деле *мыслить*, и, возможно, утверждают, что подобные компьютеры в действительности являются разумными. Или они идут еще дальше и заявляют, что эти машины могут *чувствовать* — чувствовать боль, радость, сострадание, гордость и т. п., и что они на самом деле *понимают*, что делают. То есть, как будто бы утверждается, что машины обладают *сознанием*.

Как нам понять, можно ли верить производителям? Когда мы покупаем устройство, мы, как правило, судим о его качестве лишь по полезным для нас функциям. Если устройство работает по назначению, оно нас устраивает. Если нет — его ремонтируют или меняют на новое. Чтобы проверить справедливость утверждений производителей о наличии человеческих качеств у данного устройства, мы должны, в соответствии с указанным критерием, всего лишь потребовать от устройства *поведения*, повторяющего поведение человека в отношении данных качеств. Если устройство поведет себя удовлетворительно, к производителям нет претензий, и компьютер не требует возврата для ремонта или замены.

Такая схема дает существенно операционалистский подход к рассмотрению подобных вопросов. Операционалист скажет вам, что компьютер *мыслит*, если компьютер *ведет себя* точно так же, как и человек в момент раздумий. Примем, для начала, эту операционалистскую точку зрения. Естественно, от компьютера здесь не требуется расхаживать по комнате, подобно тому, как мог бы вести себя размышляющий о чем-то человек. Еще меньше мы озабочены тем, чтобы компьютер был внешне похож на человека или напоминал на ощупь человеческое тело: эти качества не имеют отношения к назначению компьютера. То, что нас действительно интересует — его способность выдавать схожие с человеческими ответы на любой вопрос, какой нам заблагорассудится ему задать. И мы примем, что компьютер на самом деле думает (чувствует, понимает и т. д.), если его манера отвечать на наши вопросы будет неотличима от человеческой.

Этот подход очень горячо отстаивался в знаменитой статье Алана Тьюринга [1950] *Вычислительные машины и интеллект*, появившейся в 1950 году в философском журнале *Mind*. (Фамилию Тьюринг мы еще встретим позже.) В этой статье впервые была предложена идея того, что сейчас называют *тестом Тьюринга*. Тест предназначался для ответа на вопрос о том, можно ли резонно утверждать, что машина думает. Пусть утверждается, что некоторый компьютер (подобный тому, который продают производители из описания выше) в действительности думает. Для проведения теста Тьюринга компьютер вместе с человеком-добровольцем скрывают от глаз (проница-

тельной) опрашивающей [1]). Опрашивающая должна попытаться определить, где компьютер, а где человек, задавая им двоим пробные вопросы. Вопросы, а еще важнее — ответы, которые она получает, передаются в безличной форме, например, печатаются на клавиатуре и высвечиваются на экране. Единственная информация, которой будет располагать опрашивающая — это то, что она сама сможет выяснить в процессе такого сеанса вопросов и ответов. Опрашиваемый человек честно отвечает на все вопросы, пытаясь убедить женщину, что он и есть живое существо; компьютер, однако, запрограммирован таким образом, чтобы обмануть опрашивающую и убедить ее в том, что человек на самом деле он. Если в серии подобных тестов опрашивающая окажется неспособной «вычислить» компьютер никаким последовательным образом, то считается, что компьютер (или компьютерная программа, программист, разработчик и т. д.) прошел данный тест.

Можно возразить, что тест на самом деле не очень-то честный по отношению к компьютеру. Если бы роли человека и машины поменялись, и человеку нужно было бы прикидываться компьютером, определить «кто есть кто» не составило бы никакого труда: опрашивающей лишь стоило бы задать какой-нибудь очень сложный арифметический пример. Хороший компьютер тут же выдал бы правильный ответ, а человек оказался бы в замешательстве. (Здесь, однако, следует проявить осторожность. Среди людей известны «вычислительные дарования», способные в уме решать весьма нетривиальные счетные задачи с безошибочной точностью и без всяких видимых усилий. Например, сын неграмотного крестьянина Иоганн Мартин Захария Дазе [2], живший в Германии с 1824 по 1861 год, в уме перемножал любые два восьмизначных числа менее чем за минуту, а за шесть минут он перемножал два двадцатизначных числа! Такие способности не мудрено принять за результат работы компьютера. Более поздний пример (1950-е годы) — столь же исключительные вычислительные способности Александра Айткена, профессора Эдинбургского университета. Нужно, чтобы арифметическое задание опрашивающей было гораздо сложнее — например, перемножить два тридцатизначных числа за две секунды. Хороший современный компьютер запросто справится с таким упражнением.)

Итак, часть задачи программистов состояла бы в том, чтобы в некоторых вещах компьютер казался глупее, чем он есть на самом деле. Если опрашивающая задает сложный арифметический пример, подобный приведенному выше, компьютер должен притвориться, что не в силах на него ответить — иначе его немедленно изобличат! Я, правда, не думаю, что задача сделать компьютер глупее в указанном смысле является серьезной проблемой для программистов компьютеров. Главная сложность — научить компьютер отвечать на простейшие вопросы на проверку «здравого смысла», с которыми у человека вообще не будет проблем!

У конкретных вопросов такого типа есть, однако, одно слабое место. Каков бы ни был вопрос, легко придумать способ заранее научить компьютер отвечать на *данный* вопрос точно так же, как на него ответил бы человек. И тем не менее, недостаток понимания компьютером сути весьма вероятно обозначится при *продолжительном* опросе, особенно если вопросы носят нестандартный характер и требуют настоящего осмысления. Искусство опрашивающей должно включать как умение изобрести оригинальные вопросы, так и умение дополнить их позже другими вопросами на понимание таким образом, чтобы выяснить, действительно ли вопросы были усвоены. Кроме того, она может периодически подбрасывать бессмысленные вопросы (сможет ли компьютер их распознать?), или вставлять один-другой с виду бессмысленный, но на деле все-таки имеющий смысл вопрос. Например, она может спросить: «Я слышала, что

[1]) Неизбежная проблема в изложении подобного рода — дилемма между «он» и «она» в контексте, не подразумевающем никакого указания на пол. Ссылаясь на некоторый абстрактный персонаж, я далее буду использовать местоимение «он», *имея в виду* «он или она» — это, по-моему, является обычной практикой. Однако в данном случае я отдаю предпочтение женщине в роли опрашивающего, и, надеюсь, меня простят за этот единственный пример явной «половой дискриминации». Я думаю, что при определении истинно человеческих качеств женщина окажется чувствительней своего мужского двойника!

сегодня утром носорог летел вверх по Миссисипи на розовом воздушном шаре. Что Вы об этом думаете?» (Тут можно живо представить себе, как лоб компьютера покрывается капельками холодного пота — если выбрать наименее подходящую метафору.) Он может оказаться начеку и ответить: «Пожалуй, это звучит странно». Что ж, пока неплохо. Женщина: «Правда? Мой дядя как-то проделал это, причем туда и обратно, только на сероватом с полосками. Чего же тут странного?» Ясно, что без понимания компьютер скоро будет разоблачен. Отвечая на первый вопрос, он может даже ляпнуть: «Носороги не летают», — если в банках памяти удачно всплывет информация о том, что у них нет крыльев. Или ответить на второй вопрос, что носороги не бывают полосатыми. А дальше женщина может, например, подсунуть совершенно бессмысленный вопрос, заменив отдельные слова: «*под* Миссисипи», или «*внутри* розового воздушного шара» и т. п., и выяснить, хватит ли у компьютера здравого смысла, чтобы обнаружить существенное различие!

Оставим на время в стороне вопрос о том, возможно ли (а если да, то когда станет возможно) создание компьютера, который пройдет тест Тьюринга. Предположим вместо этого — исключительно для того, чтобы обсудить проблему — что такие машины уже созданы. Возникает резонный вопрос, должен ли прошедший тест компьютер *непременно* быть признан мыслящим, чувствующим, понимающим и т. д.? Этот вопрос мы рассмотрим очень скоро, а пока обсудим некоторые связанные с ним аспекты. Например такой: если производители честны во всех своих самых смелых заявлениях и их устройство есть мыслящее, чувствующее, понимающее, *сознательное* существо, то покупка устройства возлагает на нас *моральную ответственность*. Так непременно *должно быть*, если производителям можно верить. Использовать такой компьютер для наших нужд и не учитывать его переживаний было бы предосудительно. С моральной точки зрения такое использование — это то же, что и жестокое обращение с рабом. Прежде всего, мы были бы должны избегать причинить компьютеру боль, которую, по утверждениям производителей, он способен чувствовать. Выключение компьютера, возможная его продажа после того, как компьютер к нам привык, были бы сопряжены для нас с моральными проблемами. Таких проблем возникло бы великое множество, и они были бы того же сорта, что и проблемы, которые возникают у нас в отношениях с другими людьми и живыми существами. Все это стало бы для нас вопросом первостепенной важности. И крайне важной для нас (да и для административных органов!) стала бы уверенность в том, что реклама производителей типа

> Каждое мыслящее устройство прошло тщательное тестирование по Тьюрингу группой наших экспертов!

действительно является правдой.

Несмотря на очевидную абсурдность некоторых аспектов рассматриваемого вопроса (в частности, моральных), мне кажутся достаточно обоснованными доводы в пользу того, что успешно пройденный тест Тьюринга *есть* указание на присутствие мысли, интеллекта, понимания или сознания. В самом деле, на чем еще могут основываться наши убеждения в присутствии этих качеств у других людей, кроме как на беседе с ними? Строго говоря, другие критерии тоже *существуют*: выражение лица человека, движения его тела и, вообще, его действия могут оказать на нас весьма сильное влияние. Не будет ничего сверхъестественного, если (возможно, в недалеком будущем) появится робот, который сможет удачно имитировать человеческую мимику и жесты. Тогда необходимость прятать робота и человека от опрашивающей отпадет, но критерии теста, которые будут у нее в распоряжении, останутся неизменными.

Лично я готов к тому, чтобы значительно упростить тест Тьюринга. Мне кажется, что требовать от компьютера идеального подражания человеку так, чтобы стать неотличимым от него в каких-то существенных вопросах, это требовать от компьютера больше, чем надо. Мне бы хватило, чтобы наша проницательная опрашивающая по ответам на свои вопросы просто убедилась, что имеет дело с *сознательным разумом*, пусть даже чужеродным. Вот то, что реально недостижимо во всех созданных на сей день компьютерных системах. Предвижу, однако,

вероятность того, что после разоблачения компьютера у опрашивающей может возникнуть (возможно, подсознательное) нежелание приписать ему разумные качества даже тогда, когда она *способна* эти качества различить. Или наоборот, у нее может создаться впечатление «присутствия чужеродного разума», и она станет подыгрывать компьютеру, даже если «чужеродного разума» и нет. Поэтому исходный вариант теста Тьюринга гораздо предпочтительней в силу большей объективности, и ниже я обычно буду придерживаться той схемы. Присущая ей «несправедливость» по отношению к компьютеру, о которой говорилось выше (чтобы пройти тест, компьютер должен уметь все, что и человек, а человек не обязан иметь способности компьютера), не смущает сторонников теста Тьюринга, считающих этот тест точным испытанием на способность мыслить, чувствовать и т. д. Во всяком случае, многие из сторонников теста придерживаются той точки зрения, что до того, как компьютер будет способен *в действительности* пройти тест, ждать осталось недолго — скажем, до 2010 года. (По прогнозам самого Тьюринга, 30 %-ное успешное прохождение теста с опрашивающим «средних» способностей и всего с 5-минутным ограничением на продолжительность опроса могло бы быть реализовано к 2000 году.) Они уверены, что даже такая «предубежденность» не способна существенно отодвинуть эту дату!

Все вышеизложенное становится важным, коль скоро ставится вопрос по сути: дает ли операционалистская схема приемлемый набор критериев, позволяющих судить о присутствии или отсутствии мыслительных способностей у объекта? По мнению некоторых, — нет, не дает. Имитация, какой бы искусной она ни была, не должна быть с необходимостью тем же, что и оригинал. Я занимаю в этом отношении скорее промежуточную позицию. Общий принцип, к которому я склоняюсь, состоит в том, что любая, даже самая искусная, имитация всегда должна быть обнаружима достаточно тщательным тестированием. Хотя, конечно, это скорее вопрос веры (или научного оптимизма), чем доказанный факт. Таким образом, в целом я готов принять тест Тьюринга как грубо адекватный в том контексте, в котором он определяется. То есть, *если* компьютер действительно окажется способен ответить на все заданные вопросы в точности так же, как на них ответил бы человек, и тем самым последовательно и честно[2] надуть нашу проницательную опрашивающую, то *в отсутствие свидетельств об обратном* моим *предположением* было бы то, что компьютер действительно думает, чувствует и т. д. Использование мною слов «свидетельство», «действительно» и «предположение» подразумевает, что когда я говорю о мышлении, чувствах, понимании, или, в частности, *сознании*, я не отношусь к этим понятиям как к элементам общепринятой лексики, а имею в виду конкретные и объективные «вещи», присутствие или отсутствие которых в физических телах есть то, в чем мы хотели бы удостовериться. И это я считаю ключевым моментом. Пытаясь уловить присутствие данных качеств, мы делаем предположения на основании всех доступных нам свидетельств. (В принципе, точно так же действует астроном, пытаясь вычислить массу далекой звезды.)

Какие же свидетельства об обратном принимать во внимание? Наперед заданные правила установить сложно. Однако, я сразу подчеркну: тот факт, что компьютер может состоять из транзисторов и проводов, а не нейронов и кровеносных сосудов, сам по себе *не* является аргументом, который я рассматривал бы как свидетельство об обратном. Меня не покидает мысль, что когда-нибудь будет построена удовлетворительная теория сознания — удовлетворительная в смысле логической последовательности и физической приемлемости, чудесной согласованности с другим физическим знанием. Ее предсказания будут в точности соотноситься с представлениями человека об уровне и условиях существования его

[2] Я придаю особое внимание тому, что я считаю честным прохождением теста Тьюринга. Я могу, например, представить ситуацию, в которой после длинной череды поражений компьютер будет запоминать все данные ранее человеком ответы и затем выдавать их в смеси с подходящими случайными добавками. Через какое-то время у нашей уставшей опрашивающей могут кончиться нетривиальные вопросы для беседы, и она окажется обманутой компьютером — тогда я назову это «мошенничеством» с его стороны!

собственного сознания, — и такая теория может оказаться в действительности плодотворной в разрешении проблемы предполагаемого наличия сознания у нашего компьютера. Можно даже пофантазировать о «детекторе сознания», сконструированном по принципам такой теории — абсолютно надежном в случае человека, но дающем расходящиеся с тестом Тьюринга результаты в случае компьютера. Интерпретация результатов тестов Тьюринга тогда потребует особой осторожности. По моему мнению, отношение к вопросу о пригодности теста Тьюринга отчасти зависит от предположений о том, как будет развиваться наука и техника. Ниже нам еще придется вернуться к некоторым из этих рассуждений.

Искусственный интеллект

Очень большой интерес привлекают в последнее время исследования в области, называемой *искусственным интеллектом*, а часто — сокращенно — «ИИ». Целью этих исследований является научиться максимально возможно имитировать различные аспекты деятельности человеческого разума при помощи машин (как правило, электронных) и, возможно, добиться развития способностей человека в этих направлениях. Есть, по крайней мере, четыре дисциплины, которые проявляют интерес к достижениям в области ИИ. В первую очередь к ним относится *робототехника* — инженерная отрасль, которая занимается в основном индустриальными механическими устройствами, способными выполнять «интеллектуальные» операции — задачи, разнообразие и сложность которых требует вмешательства и контроля со стороны человека — причем выполнять их со скоростью и надежностью, выходящими за рамки человеческих возможностей, или в неблагоприятных условиях, где жизнь человека будет подвержена опасности. Кроме этого, как с коммерческой точки зрения, так и в целом, представляет интерес развитие *экспертных систем*, которые позволили бы закодировать самые существенные знания, относящиеся к определенным профессиям — медицинские, юридические и т. п. — в виде пакета компьютерных программ! Возможно ли, чтобы опыт и экспертные оценки специалистов этих профессий были, в самом деле, заменены такими программами? Или единственный результат этих разработок, на который можно надеяться, — это просто длинный список фактической информации с полной системой перекрестных ссылок? Вопрос о том, могут ли компьютеры демонстрировать (или симулировать) полноценную деятельность интеллекта, имеет, несомненно, весьма значительные приложения в социальной сфере. Другой областью, к которой ИИ имеет непосредственное отношение, является *психология*. Можно надеяться, что попытка смоделировать поведение человеческого мозга (равно как и мозга животного) при помощи электронных устройств — или ее поражение — позволит узнать нечто важное о высшей нервной деятельности. И, наконец, среди оптимистов бытует надежда, что по схожим причинам ИИ мог бы пролить свет на глубокие вопросы философии, дав человеку возможность проникновения в смысл понятия *разума*.

Как далеко продвинулись исследования ИИ на сегодняшний день? Я едва ли смог бы систематизированно представить здесь все достижения в этой области. В разных уголках мира существует множество активно действующих групп, с работами которых я знаком очень поверхностно. Но справедливости ради необходимо заметить, что, хотя сделано было немало, произвести что-либо, достойное называться подлинным интеллектом, до сих пор никому не удалось. Чтобы дать некоторое представление о предмете обсуждения, я для начала упомяну отдельные ранние (но даже сегодня весьма впечатляющие) достижения, а затем перейду к последним примечательным успехам в области разработки шахматных компьютеров.

Одним из первых устройств ИИ была «черепашка» Грэя В. Уолтера, созданная им в начале 1950-х годов [3], которая приводилась в движение энергией внутренних батарей и бегала по полу до тех пор, пока они почти полностью не разряжались; после чего она находила ближайшую розетку, подключалась к ней и заряжала их. Когда зарядка заканчивалась, она самостоятельно

отсоединялась и продолжала свою прогулку! В дальнейшем было придумано множество подобных механизмов (см., например, Валтц [1982]). Несколько отличное направление развития исследований представляет компьютерная программа Терри Винограда, разработанная в 1972 году, которая могла производить осмысленные действия с набором блоков, разных по форме и цвету, размещая их один над другим в разных сочетаниях и в разном порядке. Поразительно, что, несмотря на эти первые достижения, создание системы контроля даже для простой суставчатой «роботизированной руки», которая должна была в процессе перемещений избегать скопления препятствий, оказалось весьма непростой задачей, хотя стороннему наблюдателю требуемые движения и представлялись совершенно «очевидными». Такой же сложной оказалась и проблема интерпретации зрительно воспринимаемых сцен, которая в общем случае относится к области, где процедуры ИИ даже близко не подошли к реализации того, что мозг человека (и, конечно же, большинства других живых существ) способен делать «без всяких» видимых усилий.

Пожалуй, наиболее созвучна тесту Тьюринга компьютерная программа, разработанная К. М. Колби в середине 1960-х годов, которая имитирует психотерапевта — настолько успешно, что некоторые пациенты обнаружили готовность отдать ей предпочтение перед терапевтом-человеком и рассказать о себе больше, чем при общении с живым врачом. В качестве примера [4] такого общения между компьютерным «психотерапевтом» и пациентом приведем следующий отрывок из диалога (который велся посредством телетайпа):

Терапевт: Чем я могу помочь вам?

Пациент: Я не знаю...

Терапевт: Попробуйте рассказать что-нибудь об этом.

Пациент: Нет.

Терапевт: Можно узнать причину?

Пациент: Я сегодня устал и неважно себя чувствую.

Терапевт: Почему? Есть у вас на этот счет какие-нибудь предположения?

Пациент: Да... Я стар.

Терапевт: Пожалуйста, продолжайте.

Пациент: Мне кажется, что я потерял интерес ко всему происходящему.

Терапевт: Расскажите мне подробнее о ваших ощущениях.

Пациент: У меня есть кое-какие соображения по этому поводу...

Терапевт: У вас есть соображения.

Пациент: Я думаю, что мне недостает семьи.

Терапевт: Что заставляет вас переживать из-за того, что вам недостает семьи?

Пациент: В этом есть какой-то вызов...

Терапевт: Почему это так важно для вас?

Хотя этот диалог может вызвать странное ощущение понимания проблемы пациента компьютером, ничего подобного на самом деле нет: компьютер просто следует довольно простым механическим правилам. (Есть также и «обратная» ситуация, реализованная в системе, где компьютер, имитирующий страдающего шизофренией человека, дает хрестоматийные ответы и симулирует соответствующие симптомы, способные одурачить некоторых студентов-медиков и заставить их поверить в то, что с ними беседует настоящий живой пациент!)

Компьютеры, играющие в шахматы, дают, наверное, наилучший пример поведения, воспринимаемого как «разумное». В действительности, некоторые машины достигли сегодня (к 1989 году) чрезвычайно высокого мастерства игры, если сравнивать с живыми шахматистами, и приближаются к уровню мастера международного класса. (Рейтинг этих машин составляет немногим менее 2300 единиц Эло, тогда как рейтинг чемпиона мира Каспарова, для сравнения, превышает 2700.) В частности, компьютерная программа (для коммерческого микропроцессора *Fidelity Excel*), разработанная Дэном и Кейт Спраклэн, достигла показателя 2110 единиц Эло и была удостоена Шахматной федерацией США звания «Мастера». Еще больше впечатляет программа *Deep Thought*, написанная в основном Хсю (Hsiung Hsu) из университета Карнеги Меллон, рейтинг которой составляет 2500 единиц Эло и которая недавно продемонстри-

ровала замечательное достижение[3)], поделив первое место с гроссмейстером Тони Майлсом на шахматном турнире (Лонгбич, Калифорния, ноябрь 1988 года) и обыграв Бента Ларсена, что можно рассматривать, на самом деле, как первую в истории победу машины над гроссмейстером! [5] Сегодня шахматные компьютеры преуспели и в решении шахматных *задач*, с легкостью превзойдя в этом людей [6].

Шахматные машины опираются во многом на «книжные знания», помноженные на аккуратность просчета комбинаций. Стоит отметить, что машина в целом «обыгрывает» сравнимого по силе соперника в тех случаях, когда ходы необходимо делать быстро; и «проигрывает» живому противнику, если на каждый ход отпускается достаточное количество времени. Это можно понять, если принять во внимание тот факт, что компьютер принимает решения, опираясь на точные и «быстро разветвляющиеся» вычисления; тогда как преимущество живого шахматиста заключается в его способности производить «суждения», базирующиеся на сравнительно медленной сознательной деятельности по оценке ситуации. Эти человеческие суждения сводятся к тому, чтобы «отбраковать» как можно большее число возможных серьезных вариантов ходов, которые необходимо просчитывать в каждый момент; и при достаточном количестве времени на обдумывание хода такие суждения позволяют производить гораздо более глубокий анализ, чем банальное просчитывание и отбрасывание вариантов, при котором машина не использует подобные суждения. (Такая разница еще более наглядно демонстрируется в сложной восточной игре «го», где число возможностей на каждом ходу значительно больше, чем в шахматах.) Отношение между сознанием и формированием суждений будет центральным моментом в моих дальнейших рассуждениях, особенно в главе 10.

[3)] В мае 1997 года чемпион мира Г. Каспаров, рейтинг которого превышал 2800 единиц Эло, проиграл компьютерной программе *Deep Blue* со счетом 3,5 : 2,5. В октябре 2002 года матч между чемпионом мира В. Крамником и программой *Deep Fritz* закончился вничью — 4 : 4. — *Прим. ред.*

Подход к понятиям «удовольствия» и «боли» с позиций ИИ

Согласно одному из распространенных убеждений, ИИ может указать нам путь к своего рода пониманию таких категорий восприятия, как счастье, боль, голод. Возьмем, к примеру, черепашку Грэя Уолтера. Когда ее батареи садятся, ее поведение изменяется и она начинает действовать так, чтобы пополнить запас своей энергии. Здесь есть явная аналогия с тем, как человеческое существо — или любое другое животное — стало бы вести себя, ощутив голод. Похоже, мы не слишком сильно погрешим против языка, если скажем, что черепашка Грэя Уолтера была голодной, когда она действовала упомянутым образом. Некое устройство внутри нее, способное «ощущать» уровень заряда в батареях, заставляло ее переключаться в другой режим функционирования, когда заряд опускался ниже некоторой отметки. Нет причин сомневаться в том, что подобный механизм включается и в голодных животных, но с единственной разницей — изменения модели поведения в этом случае более сложны и деликатны. Вместо простого переключения с одного режима на другой здесь происходит смена *направленности* действий; и эти изменения усиливаются (до определенной степени) по мере того, как нарастает необходимость восстановить запасы энергии.

Исходя из этого, некоторые приверженцы ИИ утверждают, что такие понятия, как боль или счастье, могут быть смоделированы аналогичным образом. Давайте упростим задачу и будем рассматривать линейную шкалу «чувств», простирающуюся от крайней «боли» (отметка: −100) до абсолютного «удовольствия» (отметка: +100). Представим далее, что у нас есть устройство — какая-нибудь машина, предположительно электронная, — которая располагает средствами для регистрации собственного (условного) показателя «боль—удовольствие», который я буду называть «бу-показатель». Устройство это должно иметь определенные модели поведения и входные данные, как внутренние (типа состояния батарей), так и внешние. Идея заключается в том, что все действия машины

должны быть подчинены критерию максимизации ее бу-показателя. Факторов, влияющих на его величину, может быть множество. Мы, конечно же, можем сделать одним из них уровень заряда батарей, так, чтобы низкий уровень давал отрицательный вклад, а высокий — положительный; но могут существовать и другие факторы. Возможно, наше устройство несет на себе солнечные батареи, которые дают альтернативный источник энергии, при активации которого аккумуляторы перестают использоваться. Мы можем задать такую программу действий, при которой движение к свету будет немного увеличивать бу-показатель устройства — что оно и будет стремиться делать при отсутствии иных факторов. (Хотя, на самом деле, черепашка Грэя Уолтера, как правило, избегала света!) Ему потребуются какие-нибудь средства для выполнения вычислений, позволяющих оценивать последствия тех или иных действий в терминах величины бу-показателя. В дополнении к этому оно может уметь вводить вероятностные веса, так, чтобы в зависимости от достоверности исходных данных вычисления давали больший или меньший вклад в бу-показатель.

Помимо этого нашему устройству необходимо будет задать еще и дополнительные «цели», отличные от поддержания уровня его энергетических запасов, поскольку в противном случае мы не сможем отделить «боль» от «голода». Естественно, было бы слишком требовать от нашего механизма способности к размножению, поэтому давайте пока забудем о сексе! Но, возможно, мы могли бы имплантировать ему «желание» общения с аналогичными устройствами, приписывая таким встречам положительное значение бу-показателя. Или же мы можем заложить в него чистую «жажду знаний», когда даже простое накопление фактов об окружающем мире имело бы положительный эффект на величину бу-показателя. (Действуя из эгоистических побуждений, мы могли бы сделать так, что этот показатель увеличивался бы в результате оказания нам различных услуг — в точности, как при создании робота-слуги!) Можно было бы расценивать такой подход к назначению «целей» как искусственный, поскольку мы руководствуемся здесь разве что своими капризам. Но, в действительности, это не слишком уж отличается от способа, которым нам как индивидуумам определяются «цели» в процессе естественного отбора, где главенствующим фактором является необходимость распространять наши гены.

Предположим теперь, что мы благополучно создали наше устройство, учтя все вышеизложенные требования. Но есть ли у нас основания утверждать, что оно будет и вправду *чувствовать* удовольствие при положительном, а боль — при отрицательном значениях бу-показателя? С позиций ИИ (т. е. с операционалистской точки зрения), мы должны судить об этом просто по тому, как устройство себя ведет. Раз она действует с таким расчетом, чтобы увеличить свой бу-показатель настолько, насколько это возможно (и удерживать его на этом уровне максимально продолжительное время), и, соответственно избегать его отрицательных значений, то было бы разумным *определить* чувство удовольствия как степень положительности бу-показателя, а чувство боли — как степень его отрицательности. «Обоснованность» этого метода определения вытекает из полного сходства такого поведения с реакциями человека на удовольствие или боль. Конечно же, человеческие существа, как известно, далеко не так примитивны: иногда мы, кажется, намеренно не избавляемся от боли или избегаем некоторых удовольствий. Очевидно, что в наших действиях мы руководствуемся гораздо более сложными критериями (см. Деннетт [1978]). Но в качестве очень грубой аппроксимации можно считать, что все-таки в большинстве случаев мы стараемся избегать боли и получать удовольствие. Для операционалиста этого было бы достаточно, чтобы оправдать — в таком же приближении — идентификацию бу-показателя нашего устройства с его рейтингом по шкале «боль—удовольствие». Возможность установления подобных соответствий — одно из направлений теории ИИ.

Вопрос, который мы должны задать: правда ли, что наше устройство может по-настоящему *чувствовать* боль, если его бу-показатель отрицателен, и удовольствие в противном случае? Да и способно ли оно чувствовать хоть что-нибудь вообще? Опера-

ционалист, конечно, сказал бы «Естественно, да!»; либо отбросил бы этот вопрос как бессмысленный. Но мне представляется, что здесь *есть* серьезный и сложный вопрос, который необходимо рассмотреть. На наши действие влияет множество разнообразных факторов. Некоторые из них осознанные, как боль или удовольствие, тогда как другие мы не воспринимаем сознанием. Это наглядно иллюстрируется примером человека, касающегося раскаленной плиты. Приводится в действие механизм, который заставляет человека непроизвольно отдернуть руку еще до того, как он почувствовал боль. Вполне может оказаться, что такие спонтанные действия гораздо ближе по своей природе к реакциям нашего устройства, обусловленным его бу-показателем, чем те, которые действительно вызваны болью или удовольствием.

При описании поведения машин часто — и, обычно, в шутку — используются «человеческие» понятия: «Моя машина не хотела заводиться сегодня утром»; или «Мои часы до сих пор думают, что они идут по калифорнийскому времени»; или «Мой компьютер заявляет, что не понимает последнюю команду и не знает, что делать дальше». Конечно же, мы никоим образом не подразумеваем, что машина *действительно* может чего-либо *хотеть*, часы — что-то *думать*, а компьютер[4] — о чем бы то ни было *заявлять*, а также *понимать* или даже *знать*, что он делает. Тем не менее подобные выражения могут быть поистине информативными и способствовать нашему пониманию, при условии, что мы их будем рассматривать только в том духе, в котором будем их произносить, а не в буквальном смысле слова. Я всегда занимаю в целом аналогичную позицию по отношению к различным заявлениям сторонников ИИ о том, что сконструированные человеком устройства могут обладать характеристиками сознания — *безотносительно* от того, что под этим подразумевается! Если я согласен говорить, что черепашка Грэя Уолтера может быть голодной, то только лишь в полушутливом тоне. И если я готов использовать такие термины типа «боль» или «удовольствие», связывая их с бу-показателем некоторого устройства, как я это делал выше, то единственная причина этому заключается в том, что эти выражения облегчают мое понимание поведения устройства благодаря определенным аналогиям с моим собственным поведением и состояниями сознания. Причем здесь я ни в коем случае не подразумеваю, что эти аналогии особенно близки, или что не существует прочих — нерегистрируемых сознанием — явлений, которые влияют на мое поведение гораздо *более* схожим образом.

Я надеюсь, что читателю мое мнение достаточно ясно: я считаю, что проблема понимания свойств сознания гораздо более многогранна, чем можно извлечь непосредственно из экспериментов с ИИ. Тем не менее, я уверен в необходимости признания этой области исследований и уважительного отношения к ней. При этом я не собираюсь утверждать, будто бы достижения в задаче моделирования действительного интеллекта велики (если они вообще есть). Но нужно всегда помнить о том, что сам предмет очень «молод».

Компьютеры станут быстрее, будут обладать высокоскоростным доступом к более вместительным устройствам хранения информации, большее количество логических элементов, и научатся выполнять большее число операций параллельно. Улучшится логическая структура и техника программирования. Эти машины — носители философии ИИ — значительно и всесторонне улучшат свои возможности. Более того: сама философия отнюдь не является абсурдной по самой своей сути. Возможно, что человеческий разум может и в самом деле быть смоделирован с очень большой степенью точности при помощи электронных компьютеров — тех самых, которыми мы располагаем сегодня и принципы действия которых нам уже понятны, — но более мощных по своим характеристикам, чье появление в ближайшие годы вполне предсказуемо. Вероятно даже, что эти устройства *и вправду* будут разумными; возможно, они будут думать, чувствовать и иметь собственный интеллект. Или же, наоборот, они не будут разумными, и потребуются какие-то новые принципы, в которых мы сегодня

[4] По состоянию дел на 1989 год!

остро нуждаемся. В этом-то и заключается вопрос, от которого нельзя просто отмахнуться. Я постараюсь предоставить в ваше распоряжение факты так, как я их вижу; затем я приведу свои собственные соображения на этот счет.

Сильный ИИ и китайская комната Серла

Существует точка зрения, называемая *сильный ИИ*, которая занимает весьма радикальную позицию по этим вопросам [7]. Согласно теории сильного ИИ, не только вышеупомянутые устройства будут разумны и наделены интеллектом — свойства разума могут быть присущи логическим действиям *любого* вычислительного устройства, даже простейших из них, механических, одним из которых является, например, термостат [8]. Основная идея заключается в том, что умственная деятельность — это просто выполнение некоторой хорошо определенной последовательности операций, часто называемой *алгоритмом*. Далее я уточню это понятие. А пока нам будет достаточно определить алгоритм как своего рода вычислительную процедуру. В случае термостата алгоритм чрезвычайно прост: устройство фиксирует повышение или понижение температуры по отношению к заданной величине и размыкает или замыкает цепь, соответственно. Алгоритм, соответствующий более-менее нетривиальной деятельности головного мозга, должен быть гораздо более сложноструктурированным, но – согласно концепции сильного ИИ — это будет все же алгоритм. Он будет очень значительно отличаться от простейшего алгоритма термостата по степени сложности, но не обязательно будет иметь принципиальные отличия. Таким образом, с точки зрения сильного ИИ, существенная разница между деятельностью человеческого мозга (включая все проявления сознания) и работой термостата состоит единственно в этой самой *усложненности* (или, возможно, «структуре более высокого порядка», или «способности обращения к самому себе», или в любом другом свойстве, которое можно приписать алгоритму), имеющей место в первом случае. И, что более важно, все свойства ума — мышление, способность чувствовать, интеллект, понимание, сознание — должны рассматриваться, согласно этому подходу, просто как разные аспекты сложной деятельности; иными словами, они есть не более, чем свойства *алгоритма*, выполняемого мозгом. Достоинства любого конкретного алгоритма заключаются в его «технических характеристиках», таких как точность результатов, область применимости, экономичность и скорость выполнения. Алгоритм, нацеленный на подражание тому, что, как предполагается, действует в мозге человека, должен быть невообразимо сложным. Но если такой алгоритм для мозга существует — а это как раз то, что с уверенностью утверждают поборники идеи сильного ИИ, — то он в принципе мог бы быть запущен на компьютере. В сущности, он мог бы выполняться на любом современном компьютере общего назначения, если бы не имеющиеся ограничения по скорости и пространству для хранения данных. (Обоснование этого замечания будет дано позднее, когда мы перейдем к рассмотрению универсальной машины Тьюринга.) Предполагается, что такие ограничения будут сняты с появлением в недалеком будущем мощных быстродействующих машин. Тогда такой алгоритм, если он будет открыт, мог бы, вероятно, пройти тест Тьюринга. И как только он будет запущен, считают сторонники сильного ИИ, он будет сам по себе испытывать чувства, обладать сознанием, быть разумом.

Далеко не каждый согласится с тем, что разумные состояния и алгоритмы можно считать идентичными в указанном контексте. Наиболее остро критиковал эту точку зрения американский философ Джон Серл [1980, 1987]. Он приводил в пример ситуации, когда должным образом запрограммированный компьютер проходил упрощенную версию теста Тьюринга, и все же — он подкрепляет эти выводы очень сильными аргументами — «понимание» как свойство интеллекта полностью отсутствовало. Один из таких примеров базируется на компьютерной программе, разработанной Роджером Шенком (Шенк, Абельсон [1977]). Задачей программы была имитация понимания простых историй типа: «Мужчина вошел

в ресторан и заказал гамбургер. Когда гамбургер принесли, оказалось, что он сильно подгорел, и рассерженный мужчина выскочил из ресторана, не заплатив по счету и не оставив чаевых». В качестве второго примера можно взять другую историю: «Мужчина вошел в ресторан и заказал гамбургер. Когда его принесли, мужчина остался им очень доволен. И, покидая ресторан, он дал официанту щедрые чаевые перед тем, как заплатить по счету». Чтобы проверить «понимание» этих историй компьютером, его «попросили» определить, съел ли мужчина гамбургер в каждом отдельном случае (факт, который не был упомянут в тексте явным образом). На этот простой вопрос к таким простым историям компьютер может дать ответ, совершенно неотличимый от того, что дал бы англоговорящий человек, а именно: «нет» в первом случае и «да» — во втором. Так что в этом, *очень* узком, смысле машина уже прошла тест Тьюринга!

Вопрос, к которому мы должны далее обратиться, будет таким: действительно ли подобный положительный результат указывает на истинное понимание, демонстрируемое компьютером — или, возможно, заложенной в него программы? Как аргумент в пользу отрицательного ответа на этот вопрос, Серл предлагает свою концепцию «китайской комнаты». Он сразу же оговаривает, что истории должны рассказываться на китайском, а не на английском языке — совершенно несущественная замена — и что все команды для компьютерного алгоритма в этом конкретном случае должны быть представлены набором (английских) инструкций для работы со счетами, на которые нанесены китайские символы. Проводя мысленный эксперимент, Серл представлял, что он *сам* выполняет все манипуляции внутри запертой комнаты. Последовательность символов, описывающая истории, и вопросы к ним подаются в комнату через небольшие прорези. Никакой другой информации извне не допускается. В конце, когда все действия выполнены, последовательность, содержащая ответ, выдается из той же прорези наружу. Поскольку все эти операции есть не что иное, как составляющие процедуры выполнения алгоритма по программе Шенка, то эта последовательность должна содержать просто китайские символы, означающие «да» или «нет» и дающие корректный ответ на вопрос, который — как, собственно, и сама история — был изложен по-китайски. При этом Серл недвусмысленно дает понять, что он не знает ни слова по-китайски, и посему не имеет ни малейшего представления о содержании рассказанных историй. Тем не менее, выполнив ряд действий, составляющих алгоритм Шенка (инструкции к которому были даны ему на английском языке), он справился бы с задачей не хуже китайца, способного без труда понять эти истории. Довод Серла — и весьма сильный, по моему мнению, — заключается в том, что простое выполнение подходящего алгоритма *еще не говорит* о понимании. (Воображаемый) Серл, запертый в китайской комнате, не понимает ни на йоту, о чем идет речь в этих историях!

Против доказательства Серла был выдвинут ряд возражений. Я изложу здесь только те из них, которые — на мой взгляд — имеют серьезное значение. Прежде всего, фраза «не знает ни слова», если рассматривать ее в вышеприведенном контексте, является не вполне корректной. Понимание относится не только к отдельным словам, но и к определенным шаблонам. И при выполнении подобных алгоритмов можно в достаточной степени разобраться в структурах, которые составлены из символов, значение каждого из которых в отдельности останется непонятным. Например, китайский иероглиф, соответствующий «гамбургеру» (если он вообще существует), можно заменить на название какого-нибудь другого блюда, допустим, «чоу мейн»[5], существенно не изменив при этом содержание истории. Однако, мне все-таки кажется, что настоящий смысл историй (даже если считать такие подстановки незначительными) едва ли «дойдет» до того, кто будет просто скрупулезно выполнять шаг за шагом подобные алгоритмы.

Во-вторых, нужно всегда помнить о том, что выполнение даже сравнительно простой компьютерной программы оказывается

[5] Чоу мейн (англ. *chow mein*) — распространенное китайское блюдо на основе жареной лапши. — *Прим. ред.*

в большинстве случаев длительным и трудным процессом, если за него берется человек, манипулирующий символами. (В конце концов, именно по этой причине мы доверяем такие действия компьютерам!) Если бы Серл в самом деле выполнял указанным выше способом алгоритм Шенка, то ему для ответа на совсем простой вопрос понадобились бы дни, месяцы, а то и годы изнурительно однообразной работы — не слишком правдоподобное занятие для философа! Однако, это не представляется мне таким уж серьезным возражением, поскольку здесь мы рассматриваем вопрос *в принципе* и не касаемся технических деталей. Больше затруднений вызывает предположение о наличии компьютерной программы, способной сравниться с человеческим мозгом и, тем самым, безупречно пройти тест Тьюринга. Любая подобная программа должна быть невероятно сложной. Нетрудно вообразить, что действие такой программы, необходимое для нахождения ответа даже на сравнительно простой вопрос теста Тьюринга, состояло бы из столь большого количества шагов, что ни для одного человеческого существа выполнение соответствующего алгоритма за период, равный средней продолжительности жизни, было бы невозможным. Так ли это на самом деле — трудно сказать, не имея подобной программы в своем распоряжении [9]. Но, в любом случае, вопрос о чрезвычайной сложности (программы), по-моему, игнорировать нельзя. Понятно, что мы говорим о принципиальной стороне дела; и все же мне не кажется таким уж невероятным существование некоторой «критической» степени сложности алгоритма, которой необходимо достигнуть, чтобы алгоритм начал обладать качествами разума. Возможно, это критическое значение так велико, что ни один алгоритм, имеющий столь сложную структуру, не может быть выполнен вручную ни одним человеческим существом, как то предлагает Серл.

Сам Серл в качестве контраргумента к последнему возражению предлагает заменить фигурирующего ранее «жильца» (самого себя) китайской комнаты — целой командой не понимающих китайский язык манипуляторов символами. Чтобы сделать это число достаточно большим, он даже допускает возможность замены своей комнаты всей Индией, где все население (кроме понимающих китайский!) будет производить действия над символами. Хотя с практической точки зрения это было бы безумием, принципиально это далеко не абсурдная модель, которая не вносит существенных изменений в первоначальные выводы: те, кто манипулирует символами, по-прежнему не понимают содержание историй, вопреки утверждениям сторонников сильного ИИ о том, что простое выполнение подходящего алгоритма вызвало бы возникновение присущего интеллекту свойства «понимания». Однако, теперь это возражение оттесняется на задний план другим, кажущимся серьезнее: что, если эти индийцы более похожи на отдельные нейроны в человеческом мозгу, чем на этот мозг в целом? Никто никогда не будет ожидать от нейронов, чье возбуждение, по-видимому, является центральным механизмом умственной деятельности, чтобы они *сами* понимали, о чем думает их «хозяин» — так почему же индийцы должны понимать китайские истории? Серл парирует это возражение, указывая на явную абсурдность представления об Индии как реальной стране, понимающей некую историю, в то время как все ее население не имеет о ней ни малейшего понятия. Страна, говорит он, как и термостат или автомобиль, не «занимается» пониманием — это прерогатива индивидуумов, проживающих на ее территории.

Этот аргумент выглядит значительно слабее предыдущего. Я думаю, что доказательство Серла наиболее убедительно в случае одного исполнителя алгоритма, где мы должны ограничиться алгоритмом, чья степень сложности допускает его выполнение за время, не превышающее нормальную продолжительность человеческой жизни. Я *не* рассматриваю этот аргумент как *непреложное* свидетельство того, что не существует никакого бестелесного «понимания», ассоциируемого с процессом выполнения алгоритма людьми, чье присутствие никак не влияет на их собственное сознание. Однако, я бы скорее согласился с Серлем, что эта возможность представляется, мягко говоря, малоправдоподобной. Мне сдается, что довод Серла весьма убедителен,

хотя и не является решающим. Он с очевидностью демонстрирует, что алгоритм такой степени сложности, которой обладает компьютерная программа Шенка, не может иметь какого бы то ни было понимания выполняемых задач; также из него предположительно следует (и не более того), что ни один алгоритм, независимо от сложности его структуры, не может сам по себе воплощать настоящее понимание — вопреки утверждениям поборников сильного ИИ.

Существуют, на мой взгляд, и иные очень серьезные проблемы, связанные с сильным ИИ. Согласно этой точке зрения, единственное, что имеет значение — это алгоритм. И совершенно неважно, кто приводит его в действие: человеческий мозг, электронный компьютер, целое государство индийцев, механическое устройство из колесиков и шестеренок или система водопроводных труб. В рамках этой теории существенным для воплощения заданного «состояния разума» является сама логическая структура алгоритма, а его физическая реализация никакой роли не играет. Но, как указывает Серл, это может привести к определенной форме дуализма. Дуализм — это философское мировоззрение, апологетом которого был в высшей степени влиятельный философ и математик XVII века Рене Декарт, утверждавший, что существуют две различные субстанции: «разумная субстанция» и обычная материя. Влияют ли они друг на друга, и если да, то каким образом — это уже отдельный вопрос. Ключевое положение этой точки зрения заключается в гипотезе о том, что «разумная субстанция» не может состоять из материи обычной и способна существовать независимо от нее. «Разумная субстанция» в представлениях сильного ИИ — это логическая структура алгоритма. Как я отмечал выше, ее физическое воплощение не имеет никакого значения. Алгоритм обладает неким бесплотным существованием, никак не связанным с конкретной физической реализацией. Насколько серьезно мы должны воспринимать такой вид существования — вопрос, к которому мне придется вернуться в следующей главе. Он представляет собой часть более глобального вопроса о платонистической реальности абстрактных математических объектов. Пока же я обойду эту общую тему стороной и отмечу только, что сторонники сильного ИИ, по-видимому, принимают всерьез возможность подобного существования в случае алгоритмов, полагая, что те являются самой «сущностью» их мыслей, чувств, понимания и сознательного восприятия. В связи с этим Серл указал на примечательный в своей ироничности факт: теория сильного ИИ может привести к крайней форме дуализма — к той точке зрения, к которой сторонники сильного ИИ менее всего хотели бы иметь отношение!

Эта дилемма просматривается в рассуждениях, предложенных Дугласом Хофштадтером [1981] — убежденным сторонником сильного ИИ — в диалоге с названием *Беседа с мозгом Эйнштейна*. Хофштадтер выставляет на обозрение книгу, имеющую абсурдно большие размеры и содержащую, по его утверждению, полное описание мозга Альберта Эйнштейна. Идея такова: на любой вопрос, который кто-либо пожелал бы задать Эйнштейну, можно получить ответ в точности такой, каким был бы ответ живого Эйнштейна, если просто листать книгу и тщательно следовать всем приведенным в ней инструкциям. Конечно же, слово «просто» здесь совершенно неуместно, как то особо оговаривает сам Хофштадтер. Ведь смысл его утверждения иной: принципиально эта книга полностью эквивалентна (в операционалистском смысле теста Тьюринга) до смешного медленной «версии» настоящего Эйнштейна. Тем самым, если следовать положениям теории сильного ИИ, эта книга должна была бы думать, чувствовать, понимать и осознавать в точности так, как это делал бы сам Эйнштейн, только невероятно медленно (так что для этого «книго-Эйнштейна» внешний мир казался бы мелькающим перед ним с огромной скоростью). И естественно, что книга, представляющая из себя частную реализацию алгоритмизованной «сущности» Эйнштейна, была бы как раз-таки *самим* Эйнштейном.

Но тут возникает другая трудность. Книгу могут не открыть ни разу — или же, напротив, над ней будут корпеть многочисленные студенты и искатели истины. Как книга «поймет» разницу между этими двумя крайностями? Возможно, книгу даже

не понадобится открывать, если в ход будет пущено считывание информации при помощи рентгеновской томографии или какое-нибудь другое технологическое чудо-средство. Осознает ли Эйнштейн, что книга изучается подобным образом? Будет ли он знать о двух попытках найти с его помощью ответ на один и тот же вопрос, если он был задан дважды, разными людьми и в разное время? Или это вызовет две разделенные по времени копии *одного и то же* состояния осознания? Возможно, акт осознавания будет иметь место только в случае изменений, произошедших с книгой? В конце концов, мы обычно осознаем нечто, когда получаем о нем информацию извне, которая воздействует на наши воспоминания и, естественно, несколько изменяет состояние нашего ума. Если это так, то означает ли это, что именно (соответствующие) *изменения* алгоритмов (здесь я рассматриваю хранилище информации как часть алгоритма) должны приниматься за события, происходящие в процессе умственной деятельности — а не само *выполнение* (хотя, быть может, и оно тоже) алгоритмов? Или же «книго-Эйнштейн» способен полностью осознавать себя даже в том случае, когда его никто не будет изучать и ничто не потревожит? Хофштадтер затрагивает некоторые из этих вопросов, но на большинство из них он даже не пытается по-настоящему ответить или хотя бы подробно разобраться с ними.

Что значит «запустить алгоритм» или «реализовать его физически»? Будет ли изменение алгоритма как-нибудь отличаться от его замены на другой алгоритм? И как же все это, черт побери, связано с нашими чувствами и осознаванием?! Читатель (если только он не принадлежит к лагерю сторонников сильного ИИ) может удивиться, видя сколько времени я уделяю такой заведомо абсурдной идее. Но я-то, на самом деле, *не считаю* ее изначально абсурдной — только лишь неверной! Некоторые рассуждения, на которые опирается теория сильного ИИ, я считаю достаточно убедительными и попытаюсь обосновать свое мнение ниже. В некоторых идеях — если их модифицировать подходящим образом — есть, на мой взгляд, определенная привлекательность, которую я также постараюсь передать. Более того: как мне кажется, те самые контраргументы, которые приводит Серл, в свою очередь тоже содержат ряд серьезных головоломок и кажущихся нелепостей — хотя, в какой-то степени, я с ним и согласен!

Серл в ходе своих рассуждений неявным образом признает, что сегодняшние электронные компьютеры, снабженные значительно увеличенными быстродействием и размерами устройств хранения информации с высокой скоростью обмена данными (и, возможно, параллельным выполнением операций), вполне могли бы в обозримом будущем успешно пройти тест Тьюринга. Он готов признать утверждение сторонников сильного ИИ (и многих других «научных» точек зрения), что мы «просто конкретные экземпляры реализации некоторого числа компьютерных программ». Более того, он соглашается и с тем, что: «Конечно, наш мозг является цифровым компьютером. Поскольку всё есть цифровые компьютеры, то и мозг — тоже [10]». Серл полагает, что разница между действием человеческого мозга (который может иметь разум) и электронным компьютером (который, как он утверждает, такого свойства не имеет), когда они выполняют один и тот же алгоритм, состоит исключительно в материальной конструкции того и другого. Он заявляет — правда, не давая этому никакого обоснования — что биологические объекты (мозг) могут обладать «ментальностью» и «семантикой», которые он считает основополагающими для умственной деятельности, тогда как компьютеры — нет. Само по себе, как мне кажется, это не может указать направление развития некой полезной научной теории интеллекта. Что уж такого особенного есть в биологических системах — если не принимать в расчет их «исторический» путь развития (и того, что *мы* оказались как раз такими системами), — что могло бы выделить их в качестве объектов, которым позволено «дорасти» до ментальности или семантики? Это заявление подозрительно напоминает мне догматическое утверждение, причем не менее догматического свойства, чем утверждения сторонников сильного ИИ о том, что, просто выполняя алгоритм, можно вызвать состояние осознанного восприятия!

По-моему, Серл, как и многие другие, были введены в заблуждение компьютерщиками. А тех, в свою очередь, сбили с толку физики. (Но это не вина физиков. Даже *они* не в состоянии знать все обо всем!) Вера в то, что «все на свете является цифровыми компьютерами», кажется общераспространенной. И я намерен показать в этой книге, что это совсем не обязательно так.

«Железо» и «софт»

На компьютерном жаргоне слово «железо» используется для обозначения всех устройств и элементов, из которых состоит компьютер (печатные платы, транзисторы, провода, накопители на магнитных дисках, и т. п.), включая также полное руководство по сборке. Аналогичным образом термин «софт» относится к различным программам, которые могут выполняться на компьютере. Одним из замечательных открытий Тьюринга было то, что, по существу, любая машина с начинкой из «железа», характеризуемого определенной степенью сложности и гибкости, *эквивалентна* любой другой машине с такими параметрами. Эквивалентность двух машин (скажем, А и В) здесь должна пониматься в смысле точного соответствия действий А — при соответствующем заложенном в нее программном обеспечении — действиям В, и наоборот. Я употребляю здесь слово «точный» по отношению к конечным результатам, получающимся при введении в машины произвольных начальных данных (после того, как уже было введено преобразующее программное обеспечение), а не в смысле равенства времени, затраченного каждой машиной на получение ответа. Кроме этого, я допускаю для обеих машин возможность получения доступа к дополнительным (и, в принципе, неограниченным) внешним запасам чистых «черновиков» — магнитным пленкам, дискам, барабанам или иным носителям информации, — если какая-либо из них начинает испытывать нехватку в пространстве для хранения промежуточных результатов вычислений. Вообще говоря, разница между машинами А и В в затрачиваемом на выполнение некоторого задания времени может оказаться весьма серьезной. Вполне возможно, например, что машина А будет выполнять определенную задачу в тысячу раз быстрее, чем В. Равным образом может статься, что для другого задания время его выполнения машиной В окажется в тысячу раз меньше, чем машиной А. Более того, эти конкретные показатели могут в значительной степени зависеть от выбора используемых для конвертации программ. Но в рамках этой дискуссии нет нужды рассматривать такие практические аспекты, как способность выполнять вычисления за определенное время, поскольку наши рассуждения носят по большей части «принципиальный» характер. В следующем разделе я конкретизирую содержание тех концепций, которые затрагиваются здесь: машины А и В являют собой примеры того, что называют *универсальными машинами Тьюринга.*

В сущности, все современные общеупотребительные компьютеры — это универсальные машины Тьюринга. Тем самым все такие компьютеры будут эквивалентны друг другу в вышеупомянутом смысле: различия между ними будут заключаться единственно в программном обеспечении, при условии, что нас не волнует разница в скорости выполнения операции и возможные ограничения пространства для хранения данных. Но современные технологии сделали компьютеры способными работать так быстро и с такими огромными объемами памяти, что для большей части «повседневных» задач ни один из этих практических аспектов не накладывает серьезных ограничений на спектр решаемых такими компьютерами задач[6] — так что эта эффективная эквивалентность, введенная на теоретическом уровне, просматривается и на практике. Кажется, что технология превратила совершенно абстрактные когда-то академические дискуссии об идеальных вычислительных устройствах — в устройства реальные, и непосредственно влияющие на нашу жизнь!

Насколько я могу понять, одним из наиболее важных положений, на которых базируется философия сильного ИИ, является именно эта эквивалентность между различными физическими вычислительными

[6] См., однако, рассуждения о теории сложности и NP-задачах в конце главы 4.

устройствами. «Железо» расценивается как сравнительно (или вообще) несущественный фактор, в то время как «софт», т. е. программа или алгоритм, считается единственным жизненно важным компонентом. Однако, мне кажется, что существуют и другие, не менее важные «краеугольные камни здания сильного ИИ», которые следуют из физики. Сейчас я попытаюсь дать некоторое представление об их природе.

Что позволяет нам идентифицировать себя как личность? Может быть, в какой-то степени — сами атомы наших тел? Особое сочетания электронов, протонов и других частиц, из которых состоят эти атомы? Есть, по крайней мере, два возражения против этого предположения. Во-первых, вещество тела любого живого существа претерпевает постоянные изменения и обновления. Это справедливо, в частности, для клеток головного мозга, несмотря на то, что после рождения новые клетки уже не образуются. Абсолютное большинство атомов в каждой живой клетке (включая все клетки мозга) — и, конечно же, практически все ткани нашего тела — замещаются новыми по много раз с момента рождения.

Второе возражение приходит из квантовой физики — и, по странной иронии, находится, строго говоря, в прямом противоречии с первым! Согласно квантовой механике (и мы узнаем об этом больше в главе 6, с. 246) любые два электрона должны быть с необходимостью одинаковыми; и то же самое справедливо в отношении двух произвольно взятых протонов или пары любых других частиц, относящихся к одному типу. То, что подразумевается под этим, отнюдь не ограничивается утверждением об их неразличимости — оно значительно сильнее. Если пришлось бы поменять между собой электрон в человеческом мозге и электрон в кирпиче, то состояние системы осталось бы *в точности тем же самым* [11], что и до этого — тем же самым, а не просто неотличимым! Аналогичное правило справедливо и для протонов, и для других разновидностей частиц, а также для целых атомов, молекул и т. п. Если весь материал человеческого тела заместить соответствующими частицами кирпичей из его дома, то, в буквальном смысле, вообще ничего не изменится. То, что отличает человека от своего дома — это то, в какую *структуру* организованы составляющие его тела, а не индивидуальные свойства этих составляющих.

Можно привести аналогию из повседневной жизни, не имеющую отношения к квантовой механике, которая бросилась мне в глаза, пока я набирал эти строки, имея в своем распоряжении один из плодов информационной технологии — текстовый редактор. Если я хочу изменить слово, скажем, «болт» на «борт», то могу сделать это просто заменив букву «л» буквой «р»; или же я могу вместо этого напечатать все слово заново. Выбрав последний вариант, я встану перед вопросом: а та ли это теперь буква «б», что была ранее, или я заменил ее идентичной? А как насчет «т»? Даже если я решу просто поменять букву «л» на «р», а не перебивать все слово заново — будет момент, как раз между удалением «л» и появлением «р», когда пустое место «схлопывается» и по всему тексту сверху вниз пройдет волна перестановок, при которых пересчитывается расположение всех букв, включая «т» — а затем пере-пересчитывается еще раз при вставке на то же место «р». (Ох уж эта дешевизна бездумных вычислений в наши дни!) В любом случае, все буквы, которые я вижу на экране, есть не более чем разрывы на пути следования электронного луча в процессе сканирования всего экрана, происходящего шестьдесят раз в секунду. Если я возьму произвольную букву и заменю ее на такую же — сохранится ли при этом исходное состояние *точно таким же* или оно будет только лишь неотличимо? Попытка провести смысловое разделение между двумя этими определениями нового состояния (т. е. между «только лишь неотличимое» и «точно такое же») кажется несерьезной. По крайней мере, коль скоро замещающая буква является идентичной, возникает желание назвать это состояние таким же. И то же самое верно и для квантовой механики одинаковых частиц. Поменять одну из частиц на другую, эквивалентную — все равно, что не поменять ничего. Состояние при этом должно считаться *тем же самым*, что и в начале. (Однако, как станет ясно в главе 6, подобное различие *не так уж* тривиально в контексте квантовой механики.)

Рассуждения, сделанные выше по поводу непрерывного обновления атомов человеческого тела, надо рассматривать скорее в рамках классической физики, нежели квантовой. В этих рассуждениях используется терминология, которая неявно подразумевает возможность индивидуального существования каждого атома. На этом уровне описания классическая физика вполне адекватна и мы не слишком погрешим против истины, если будем рассматривать атомы в качестве отдельных объектов. При условии, что атомы достаточно хорошо отделены друг от друга в процессе движения, *можно* было бы говорить об их индивидуальном существовании, поскольку каждый атом допускает в этом случае непрерывное наблюдение за собой. С точки зрения квантовой механики говорить об индивидуальности атомов можно только ради удобства описания, однако на рассматриваемом уровне это вполне допустимо.

Давайте примем, что индивидуальность человека никак не связана с индивидуальностью, которую можно было бы постараться приписать его материальной основе. Вместо этого она должна определяться своего рода *конфигурацией* составляющих элементов этой основы — их пространственной или, допустим, пространственно-временной структурой. (Подробно об этом — далее.) Но сторонники сильного ИИ идут еще дальше. Если информационное содержание такой конфигурации перевести в другую форму, из которой затем можно было бы полностью восстановить оригинал, то, согласно их утверждению, индивидуальность человека осталась бы неизменной. Это похоже на ситуацию с последовательностью букв, которую я только что напечатал и теперь вижу на дисплее моего текстового редактора. Если я уберу их с экрана, то они, тем не менее, сохранятся записанными в виде определенных крошечных изменений электрического заряда, в конфигурации, геометрически никак не соотносящейся с буквами, которые я минуту назад напечатал. И все же в любой момент я могу вернуть их на экран — и вот они, пожалуйста, точь-в-точь такие же, словно и не было никаких преобразований. Если я захочу сохранить написанное, то я могу перевести информацию о последовательности букв в некоторую конфигурацию намагниченных доменов на диске, который я затем выну и выключу машину, аннулируя тем самым все (соответствующие) крошечные изменения заряда в ячейках ее памяти. Тогда завтра я смогу снова вставить диск, восстановить эти смещения и отобразить последовательность букв на экране так, как будто ничего и не случилось. Приверженцам теории сильного ИИ «ясно», что аналогичным образом можно обращаться и с личностью человека. Как и в случае с буквами у меня на экране, скажут они, человеческая индивидуальность ничего не потеряла бы — собственно, с ней вообще ничего бы не произошло, — если ее физическую форму перевести во что-нибудь совершенно иное, скажем, в поля намагниченности железного бруска. Они, кажется, даже готовы поспорить, что сознательное восприятие человека сохранилось бы и в то время, пока «информация» о нем пребывает в другой форме. При таком подходе «человеческое сознание» должно рассматриваться, по сути, как набор программ — «софта», — а его конкретное воплощение в виде материального человеческого существа — как действия этих программ, осуществляемые «железной начинкой» его тела и мозга.

Основанием для подобных заявлений служит, вероятно, убежденность в том, что какую бы материальную форму не принимало «железо» — пусть это будет, например, какое-нибудь электронное устройство, — ему можно будет всегда «задать» вопрос-программу (в духе теста Тьюринга), и ответ на него, в предположении о способности «железа» адекватно вычислять ответы на эти вопросы, будет неотличим от ответа человека, данного им в нормальном психическом состоянии. («Как вы чувствуете себя сегодня утром?» — «О, вполне сносно, хотя мне немного докучает легкая головная боль». — «Значит, вы не чувствуете... э-э... ну, чего-нибудь необычного, связанного с вашей личностью... ничего такого?» — «Нет. А почему вы спрашиваете об этом? Довольно странный, знаете ли, вопрос...» — «То есть вы чувствуете себя тем же самым человеком, что и вчера?» — «Ну конечно!»)

Идея, которую часто обсуждают в связи с этим, носит в фантастической литературе название *телепортационной машины*[12]. Предполагается использовать ее для транспортировки, допустим, с одной планеты на другую; но будет ли она работать именно таким образом — это как раз и является предметом обсуждения. Вместо того, чтобы перемещаться «обычным» путем — на космическом корабле, — гипотетический путешественник подвергается сканированию с макушки до пят, при котором со всей возможной аккуратностью фиксируется положение и характеристики каждого атома в его теле. Затем вся эта информация передается со скоростью света при помощи любого подходящего электромагнитного сигнала на ту планету, где он хотел бы оказаться. Там эта информация собирается воедино и используется в качестве инструкций для создания точной копии путешественника, со всеми его воспоминаниями, устремлениями, надеждами и самыми глубокими чувствами. По крайней мере, так это должно выглядеть на практике: все детали состояния мозга подробно записываются, затем передаются, и по этим данным происходит реконструирование. Если предположить, что все произошло так, как надо, то оригинал можно «безболезненно» уничтожить. В таком случае возникает вопрос: является ли такой механизм *настоящим* путешествием с одного места на другое — или же это просто создание дубликата, сопровождающееся убийством оригинала? Будете ли вы готовы воспользоваться таким способом «путешествия» при условии, что он подтвердит свою стопроцентную надежность? Если телепортация *не является* путешествием, то в чем же заключается *принципиальная* разница между ней и простым переходом из одной комнаты в другую? А в последнем случае — разве не определяют атомы в один момент времени информацию об их положении в последующие моменты? В конце концов, мы видели, что сохранять «индивидуальность» какого бы то ни было атома — нецелесообразно. Вопрос об индивидуальных характеристиках атома вообще не имеет смысла. Разве произвольная движущаяся структура из атомов не представляет собой своего рода волну информации, распространяющуюся между точками пространства? Тогда есть ли существенная разница между распространением волн, несущих информацию о переходящем из комнаты в комнату человеке, — и тех, что посылаются устройством телепортации?

Допустим, что телепортация действительно «работает» в том смысле, что «сознание» путешественника на самом деле просыпается в его двойнике, находящемся на далекой планете. Что тогда произойдет в том случае, если мы, в нарушение правил игры, не уничтожим оригинал путешественника? Будет ли его «сознание» одновременно в двух разных местах? (Попытайтесь представить свою реакцию на следующее заявление: «Ах, дорогой, похоже, суспензия, которую мы дали тебе перед посадкой в Телепортатор, испортилась раньше срока? Да, вышло не очень удачно, хотя это не так страшно. В любом случае, тебе, наверное, будет приятно услышать, что другой ты — ну-у, то есть, конечно, *настоящий* ты — прибыл на Венеру в целости и сохранности, поэтому мы можем... э-э... избавиться от тебя здесь — нет, я имею виду... ну, от ненужной больше копии. Разумеется, это пройдет совершенно безболезненно».) Возникает парадоксальная ситуация. Существуют ли в физике законы, делающие телепортацию *принципиально* невозможной? С другой стороны, возможно, там нет никаких абсолютных запретов на такую «передачу» человека и его сознания, но сам принцип «копирования» предполагает неизбежное уничтожение оригинала? Может быть, сохранение двух дееспособных копий запрещено в принципе? Хотя эти рассуждения носят отстраненный характер, я все же верю, что из них можно извлечь кое-какие полезные сведения о физической природе сознания и индивидуальности. Я вижу в них явное указание на ту существенную роль, которую играет квантовая механика в понимании явлений умственной деятельности. Но я слишком забегаю вперед. К этой теме необходимо будет вернуться после того, как мы изучим структуру квантовой теории в главе 6 (см. с. 240).

Давайте посмотрим, какое отношение имеет теория сильного ИИ к вопросу о телепортации. Мы предположим, что где-то между двумя планетами располагается ре-

трансляционная станция, на которой полученная информация некоторое время хранится перед тем, как быть отправленной к месту своего назначения. Для удобства эта информация записывается не в человеческой форме, а в каком-нибудь электронном или магнитном устройстве. Будет ли человеческое «сознание» присутствовать в этом устройстве? Приверженцы сильного ИИ постарались бы убедить вас в том, что это будет именно так. Ведь в конечном счете, сказали бы они вам, на любой вопрос, который мы решили бы задать путешественнику, могло бы, в принципе, ответить и это устройство — если «просто» сымитировать соответствующую функцию его мозга. Устройство располагало бы всей необходимой информацией, и дело стало бы только за вычислениями. А если устройство отвечает на вопросы в точности также, как если бы это был путешественник, то (с точки зрения теста Тьюринга!) оно им и *является*. В качестве основы для такого вывода здесь опять выступает известное утверждение сторонников сильного ИИ: для явлений, связанных с умственной деятельностью, «железо» не имеет никакого значения. Это утверждение кажется мне неправомочным. Оно, в свою очередь, основывается на представлении о мозге (или разуме) как о цифровом компьютере. И подразумевает, что нет каких-то особых физических процессов, приводящихся в действие, когда человек думает, которые могли бы требовать для своей реализации ту конкретную физическую (биологическую, химическую) структуру, которой обладает мозг.

Естественно, проповедники сильного ИИ будут настаивать на том, что единственное предположение, которое при этом вводится, касается универсальной возможности численного моделирования любого физического процесса. Я более чем уверен, что подавляющее большинство физиков, опираясь на современное состояние физической науки, сочло бы такое предположение совершенно оправданным. В следующих главах я представлю свои собственные доводы в пользу противоположной точки зрения (а также подготовлю почву, чтобы объяснить, почему я думаю, что делается некое предположение). Но давайте на мгновение примем (широко распространенную) точку зрения, согласно которой все относящиеся к предмету дискуссии физические процессы *допускают* численное моделирование. Тогда единственным (если не принимать во внимание вопросы о времени и ресурсах, затраченных на вычисления) реальным предположением будет следующее «операционалистское» предположение: если нечто *действует* в точности, как существо, обладающее осознанным восприятием, то мы должны считать, что оно себя этим существом и «чувствует».

Точка зрения теории сильного ИИ состоит в том, что, рассматривая «только» вопрос, относящийся к «железу», любые физические процессы, имеющие отношение к работе мозга, в обязательном порядке могут быть промоделированы с помощью соответствующего преобразующего «софта». Если мы принимаем операционалистскую точку зрения, то тогда этот вопрос будет состоять в эквивалентности универсальных машин Тьюринга, в том, что такие машины способны выполнять любой алгоритм, — а также в справедливости предположения об алгоритмической природе деятельности мозга. И теперь самое время коснуться этих интригующих и важных понятий более подробно.

Примечания

1. См., например, работы Гарднера [1958], Грэгори [1981] и содержащиеся там ссылки.
2. См., например, работу Резникова и Уэллса [1984]. Ознакомиться с классическими обзорными работами по «чудесам» вычислений можно у Рауза Болла [1982] и Смита [1983].
3. См. работы Грэгори [1981] и Уолтера [1953].
4. Этот пример взят из Дельбрюка [1986].
5. Смотри статьи О'Коннелла [1988] и Кина [1988]. За дальнейшей информацией по компьютерным шахматам я отсылаю читателя к Леви [1984].

6. Конечно же, сложность большинства шахматных задач рассчитывалась на *людей*. Возможно, было бы не так уж трудно придумать шахматную задачу, не очень сложную для человеческого существа, но такую, что современные шахматные компьютеры не смогли бы решить и за тысячу лет. (Принцип подобной задачи достаточно очевиден: она должна состоять из очень большого числа ходов. Известны задачи, требующие для решения порядка 200 ходов — более чем достаточно!)
7. Везде в этой книге я использую термин Серла «сильный ИИ» для обозначения этой радикальной точки зрения просто чтобы быть точным. Слово «функционализм» часто применяется по отношению к такому же по сути воззрению, но, наверное, не всегда корректно. Этой точки зрения придерживаются Мински [1968], Фодор [1983], Хофштадтер [1979], Моравец [1989].
8. См. работу Серла [1987] в качестве примера такого утверждения.
9. Дуглас Хофштадтер в своей критике оригинальной работы Серла (так, как она перепечатана в *The Mind's I*) возражает, что ни одно человеческое существо не в состоянии «разобраться» в полном описании разума другого человека из-за большой сложности. И это действительно так! Но мне кажется, что идея не в этом. Ведь выполнить нужно будет только ту часть алгоритма, которая должна отвечать какому-то одному мыслительному процессу. Таким могло бы оказаться некое мгновенное «осознание» при ответе на вопрос теста Тьюринга, или даже что-нибудь еще более простое. А кто сказал, что подобное действие с необходимостью потребовало бы выполнения алгоритма невообразимой сложности?
10. См. статью Серла [1980], которая была опубликована в книге Хофштадтера и Деннетта [1981].
11. Некоторые читатели, сведущие в этом вопросе, могли бы возмутиться из-за некоторой разницы в знаках. Но даже это (спорное) различие пропадет, если мы при замене повернем один из электронов на 360 градусов! (Пояснения можно найти на с. 246, глава 6.)
12. См. вступление к книге Хофштадтера и Деннетта [1981].

Глава 2

Алгоритмы и машины Тьюринга

Основы алгоритмов

Как точно определить понятие алгоритма, или машины Тьюринга, или универсальной машины Тьюринга? Почему эти понятия играют одну из главных ролей в современном представлении о «мыслящем устройстве»? Есть ли какие-нибудь абсолютные ограничения на принципиальные возможности использования алгоритмов? Для того чтобы ответить на эти вопросы, нам придется разобраться в деталях, что представляют собой алгоритм и машины Тьюринга.

В дальнейших рассуждениях я буду иногда прибегать к математическим выражениям. Вероятно, некоторых читателей эти выкладки напугают и даже заставят отложить книгу в сторону. Если вы как раз такой читатель, то я прошу вашего снисхождения и рекомендую вам последовать совету, данному мной в *Обращении к читателю* на с. 26! Доказательства, которые здесь встретятся, не потребуют владения математическим аппаратом, выходящим за пределы школьного курса, но чтобы в них детально разобраться, все же понадобятся интеллектуальные усилия. На самом деле, большинство рассуждений изложено весьма подробно, и если внимательно им следовать, можно добиться глубокого понимания. Однако, даже беглый просмотр доказательств позволяет ухватить основную идею. С другой стороны, если вы являетесь экспертом в этой области, то я опять вынужден принести свои извинения. Но я осмелюсь предположить, что даже в этом случае вам будет небесполезно ознакомиться с моими рассуждениями, в которых почти наверняка найдется что-то интересное и для вас.

Слово «алгоритм» происходит от имени персидского математика IX века Абу Джафара Мухаммеда ибн Мусы аль-Хорезми, написавшего около 825 года н. э. руководство по математике *Kitab al-jabr wa'l-muqabala*, которое оказало значительное влияние на математическую мысль того времени. Современное написание «алгоритм», пришедшее на смену более раннему и точному «алгоризм», своим происхождением обязано, скорее всего, ассоциации со словом «арифметика»[1]. (Примечательно, что и слово «алгебра» происходит от арабского *al-jabr*, фигурирующего в названии вышеупомянутой книги.)

Примеры алгоритмов были, однако, известны задолго до появления книги аль-Хорезми. Один из наиболее известных — алгоритм Евклида — процедура отыскания наибольшего общего делителя двух чисел, восходит к античности (примерно 300 лет до н. э.). Давайте посмотрим, как он работает. Возьмем для определенности два числа, скажем, 1365 и 3654. Наибольшим общим делителем двух чисел называется самое большое натуральное число, на которое делится каждое из этих чисел без остатка. Алгоритм Евклида состоит в следующем. Мы берем одно из этих чисел, делим его

[1] Автор имеет в виду созвучность английских слов *algorithm* и *arithmetic*. — *Прим. ред.*

на другое и вычисляем остаток: так как 1365 входит дважды в 3654, в остатке получается 3654 − 2 × 1365 = 924. Далее мы заменяем наши два исходные числа делителем (1365) и полученным остатком (924), соответственно, производим с этой парой ту же самую операцию и получаем новый остаток: 1365 − 924 = 441. Для новой пары чисел — а именно, 924 и 441, — получаем остаток 42. Эту процедуру надо повторять до тех пор, пока очередная пара чисел не поделится нацело. Выпишем эту последовательность:

3654 : 1365	дает в остатке	924
1365 : 924	дает в остатке	441
924 : 441	дает в остатке	42
441 : 42	дает в остатке	21
42 : 21	дает в остатке	0.

Последнее число, на которое мы делим, а именно 21, и есть искомый наибольший общий делитель.

Алгоритм Евклида является *систематической процедурой*, которая позволяет найти этот делитель. Мы только что применили эту процедуру к двум конкретным числам, но она работает и в самом общем случае с произвольными числами. Для очень больших чисел эта процедура может занять много времени, и будет выполняться тем дольше, чем больше сами числа. Но в каждом конкретном случае выполнение процедуры в конце концов заканчивается, приводя за конечное число шагов к вполне определенному ответу. На каждом этапе мы точно представляем себе действие, которое должно быть выполнено, и точно знаем, когда получен окончательный результат. Более того, всю процедуру можно описать *конечным* числом терминов, несмотря на то, что она может применяться к любым, сколь угодно большим натуральным числам. («Натуральными числами» называются неотрицательные [1] целые числа 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, … .) На самом деле нетрудно изобразить (конечную) блок-схему, описывающую логическую последовательность операций алгоритма Евклида (рис. 2.1).

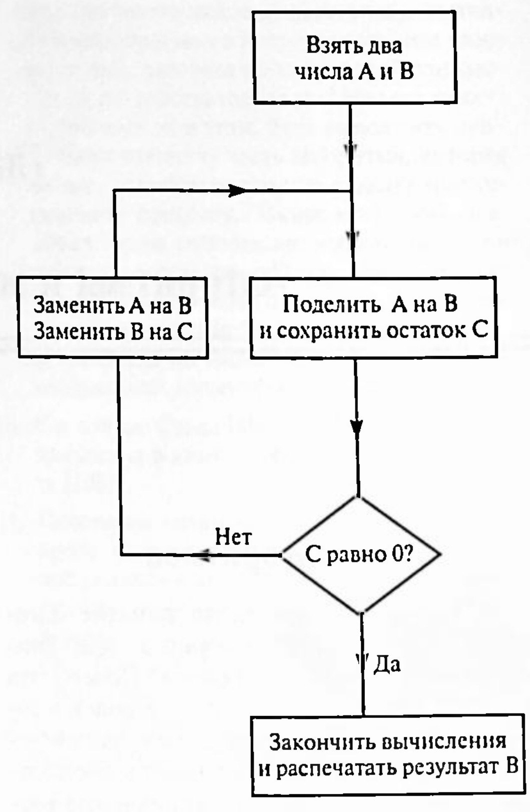


Рис. 2.1

Нужно заметить, что на схеме эта процедура не до конца разбита на простейшие составляющие, поскольку мы неявным образом предположили, что нам уже «известно», как выполнять необходимую базовую операцию получения остатка от деления двух произвольных натуральных чисел A и B. Эта операция, в свою очередь, также алгоритмична и выполняется при помощи хорошо знакомой нам со школы процедуры деления. Эта процедура, на самом деле, сложнее, чем все остальные части алгоритма Евклида, но и она может быть представлена в виде блок-схемы. Основное затруднение здесь возникает из-за использования привычной «десятичной» записи натуральных чисел, что вынуждает нас выписывать все таблицы умножения, учитывать перенос и т. п. Если бы для представления некоторого числа n мы использовали последовательность из n каких-нибудь одинаковых знаков, например, пяти звездочек (*****) для обозначения пятерки, то определение остатка свелось бы к совершенно элементарной алгоритмической операции. Для того чтобы получить остаток от деления A на B, достаточно просто убирать из записи числа A последовательность знаков, представляющих B, до тех пор, пока на некотором

этапе оставшееся число знаков в записи A не станет недостаточным для выполнения следующего шага. Эта последовательность знаков и даст требуемый ответ. Например, желая получить остаток от деления 17 на 5, мы просто будем последовательно удалять ***** из *****************, как это показано ниже:

****
****
****
**,

и в результате получим, очевидно, 2, так как следующее удаление уже станет невозможно.

Блок-схема изложенного выше процесса нахождения остатка от деления путем последовательных вычитаний приведена на рис. 2.2. Чтобы придать блок-схеме алгоритма Евклида завершенный вид, мы должны подставить схему отыскания остатка в соответствующий блок справа в центре предыдущей схемы. Такая подстановка одного алгоритма в другой — распространенная в компьютерном программировании процедура. Алгоритм вычисления остатка, изображенный на рис. 2.2, служит примером *подпрограммы*, иначе говоря, это алгоритм (как правило, уже известный), вызываемый и используемый по мере надобности в ходе выполнения основного алгоритма.

Безусловно, обозначение числа n просто набором из n звездочек чрезвычайно неэффективно, когда речь заходит о больших числах. Именно поэтому обычно используют более компактную запись, например, стандартную (десятичную) систему. Однако оставим в стороне *эффективность* операций и обозначений и уделим все внимание вопросу о том, какие операции *в принципе* могут выполняться алгоритмически. Действие, которое поддается алгоритмизации в одной записи, сохранит это свойство и в любой другой. Эти два случая различаются только техническими нюансами и сложностью выполнения алгоритма.

Алгоритм Евклида — это лишь одна из многих, часто классических, алгоритмических процедур, встречающихся в математике повсеместно. Но, вероятно, не лишним будет отметить, что, несмотря на значительный исторический возраст отдельных

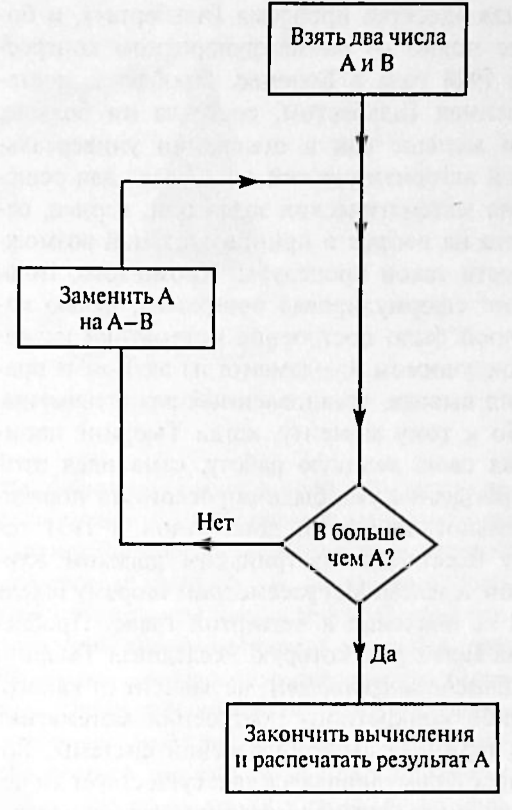


Рис. 2.2

алгоритмов, точная формулировка *универсального определения алгоритма* появилась только в *двадцатом* веке. В 1930-х годах было предложено несколько альтернативных формулировок этого понятия, из которых наиболее емкая и убедительная — и, к тому же, наиболее значимая в историческом плане — опирается на понятие *машины Тьюринга*. Поэтому нам будет полезно рассмотреть некоторые свойства этих «машин».

Прежде всего следует помнить, что «машина» Тьюринга принадлежит области «абстрактной математики» и ни в коем случае не является физическим объектом. Это понятие было введено в 1935–1936 годах английским математиком и кибернетиком Аланом Тьюрингом, внесшим огромный новаторский вклад в развитие компьютерной науки (Тьюринг [1937]). Тьюринг рассматривал задачу весьма общего характера (известную как проблема алгоритмической разрешимости), которая была поставлена великим немецким математиком Давидом Гильбертом частично в 1900 году на Парижском Конгрессе математиков (так называ-

емая «десятая проблема Гильберта»), и более полно — на международном конгрессе 1928 года в Болонье. Проблема, поставленная Гильбертом, состояла ни больше, ни меньше как в отыскании универсальной алгоритмической процедуры для решения математических задач или, вернее, ответа на вопрос о принципиальной возможности такой процедуры. Кроме того, Гильберт сформулировал программу, целью которой было построение математики на несокрушимом фундаменте из аксиом и правил вывода, установленных раз и навсегда. Но к тому моменту, когда Тьюринг написал свою великую работу, сама идея этой программы уже была опровергнута поразительной теоремой, доказанной в 1931 году блестящим австрийским логиком Куртом Геделем. Мы рассмотрим теорему Геделя и ее значение в четвертой главе. Проблема Гильберта, которую исследовал Тьюринг (*Entscheidungsproblem*), не зависит от какого-либо конкретного построения математики в терминах аксиоматической системы. Вопрос формулировался так: существует ли некая универсальная механическая процедура, позволяющая, в принципе, решить все математические задачи (из некоторого вполне определенного класса) одну за другой?

Трудность с ответом на этот вопрос была связана отчасти с определением смысла «механической процедуры» — это понятие выходило за рамки стандартных математических идей того времени. Чтобы как-то ее преодолеть, Тьюринг постарался представить, как можно было бы формализовать понятие «машина» путем расчленения ее действий на элементарные операции. Вполне вероятно, что в качестве примера «машины», помимо прочего, Тьюринг рассматривал и человеческий мозг, тем самым относя к «механическим процедурам» все действия, которые математики выполняют, размышляя над решением математических задач.

Хотя такой взгляд на процесс мышления оказался весьма полезным при разработке Тьюрингом его в высшей степени важной теории, нам совершенно необязательно его придерживаться. Действительно, дав точное определение механической процедуры, Тьюринг тем самым показал, что существуют совершенно четко определенные математические операции, которые никак не могут называться механическими в общепринятом смысле слова. Можно, наверное, усмотреть некую иронию в том, что эта сторона работы Тьюринга позволяет нам теперь косвенным образом выявить его собственную точку зрения на природу мышления. Однако, нас это пока занимать не будет. Прежде всего нам необходимо выяснить, в чем же, собственно, заключается теория Тьюринга.

Концепция Тьюринга

Попробуем представить себе устройство, предназначенное для выполнения некоторой (конечноопределенной) вычислительной процедуры. Каким могло бы быть такое устройство в общем случае? Мы должны быть готовы к некоторой идеализации и не должны обращать внимания на практические аспекты — мы на самом деле рассматриваем математическую идеализацию «машины». Нам нужно устройство, способное принимать дискретное множество различных возможных состояний, число которых *конечно* (хотя и может быть очень большим). Мы назовем их *внутренними* состояниями устройства. Однако мы не хотим, чтобы объем выполняемых на этом устройстве вычислений был принципиально ограничен. Вспомним описанный выше алгоритм Евклида. В принципе, не существует предельной величины числа, после которой алгоритм перестает работать. Этот алгоритм, или некая общая *вычислительная процедура*, будет тем же самым независимо от того, сколь велики числа, к которым он применяется. Естественно, для очень больших чисел выполнение процедуры может занять много времени и может потребоваться огромное количество «черновиков» для выполнения пошаговых вычислений. Но сам по себе *алгоритм* останется тем же *конечным* набором инструкций, сколь бы большими ни были эти числа.

Значит, несмотря на конечность числа внутренних состояний, наше устройство должно быть приспособлено для работы с входными данными неограниченного объема. Более того, устройство должно иметь возможность использовать внешнюю память

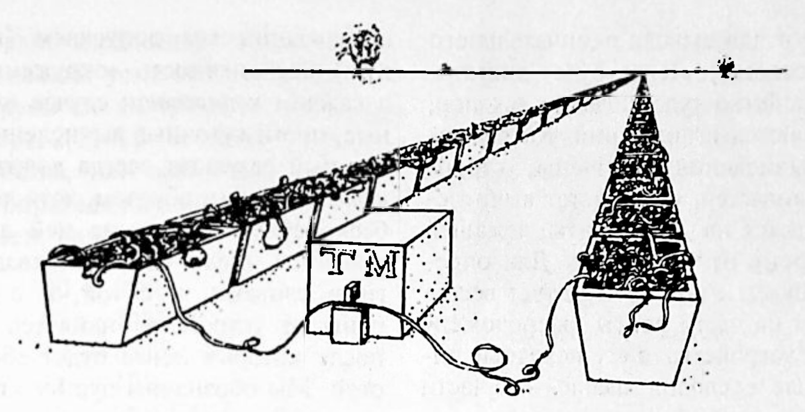


Рис. 2.3. Точная машина Тьюринга требует бесконечной ленты!

неограниченного объема (наши «черновики») для хранения данных, необходимых для вычислений, а также уметь выдавать окончательное решение любого размера. Поскольку наше устройство имеет только конечное число различных внутренних состояний, мы не можем ожидать, что оно будет «хранить внутри себя» все внешние данные, равно как и результаты своих промежуточных вычислений. Напротив, оно должно обращаться только к тем данным и полученным результатам, с которыми оно работает непосредственно в настоящий момент, и уметь производить над ними требуемые (опять же, в данный момент) операции. Далее, устройство записывает результаты этих операций — возможно, в отведенной для этого внешней памяти — и переходит к следующему шагу. Именно неограниченные объемы входных данных, вычислений и окончательного результата говорят о том, что мы имеем дело с идеализированным математическим объектом, который не может быть реализован на практике (рис. 2.3). Но подобная идеализация является очень важной. Чудеса современных компьютерных технологий позволяют создавать электронные устройства хранения информации, которые мы можем рассматривать как неограниченные в приложении к большинству практических задач.

На самом деле память устройства, которая выше была названа «внешней», можно рассматривать как внутренний компонент современного компьютера. Но это уже технические детали — рассматривать часть объема для хранения информации как внутреннюю или внешнюю по отношению к устройству. *Одним* из способов проводить такое деление между «устройством» и «внешней» частью могло бы стать использование понятий *аппаратного* (*hardware*) и *программного* (*software*) обеспечения вычислений. В этой терминологии внутренняя часть могла бы соответствовать *аппаратному обеспечению* (*hardware*), тогда как внешняя — *программному обеспечению* (*software*). Я не буду жестко придерживаться именно этой классификации, однако, какую бы точку зрения мы не заняли, не вызывает сомнений, что идеализация Тьюринга достаточно точно аппроксимируется современными электронными компьютерами.

Тьюринг представлял внешние данные и объем для хранения информации в виде «ленты» с нанесенными на нее метками. Устройство по мере необходимости могло обращаться к этой ленте, «считывать» с нее информацию и перемещать ее вперед или назад в ходе выполнения операций. Помимо этого, устройство могло ставить новые метки на ленту и стирать с нее старые, что позволяло использовать *одну и ту же* ленту и как внешнюю память (то есть «черновик»), и как источник входных данных. На самом деле, не стоило бы проводить явное различие между этими двумя понятиями, поскольку во многих операциях промежуточные результаты вычислений могут играть роль новых исходных данных. Вспомним, что при использовании алгоритма Евклида мы раз за разом замещали исходные числа (A и B) результатами, полученными на разных этапах вычислений. Сходным образом та же самая лента может быть

использована и для вывода окончательного результата («ответа»). Лента будет двигаться через устройство туда-сюда до тех пор, пока выполняются вычисления. Когда, наконец, все вычисления закончены, устройство останавливается, и результат вычислений отображается на части ленты, лежащей по одну сторону от устройства. Для определенности будем считать, что ответ всегда записывается на части ленты, расположенной слева от устройства, а все исходные числовые данные и условия задачи — на части ленты, расположенной справа от него.

Меня всегда несколько смущало представление о конечном устройстве, которое двигает потенциально бесконечную ленту вперед и назад. Неважно, насколько легок материал ленты — сдвинуть *бесконечную* ленту все-таки будет трудно! Вместо этого я предпочитаю представлять себе эту ленту как некое окружение, по которому может перемещаться наше конечное устройство. (Конечно же, в современных электронных устройствах ни «лента», ни само «устройство» не должны в обычном смысле физически «перемещаться», но представление о таком «движении» позволяет достичь известной наглядности.) При таком подходе устройство получает все входные данные из этого окружения, использует его в качестве «черновика» и, наконец, записывает в него конечный результат.

В представлении Тьюринга «лента» состоит из бесконечной в обоих направлениях линейной последовательности квадратов. Каждый квадрат либо пуст, либо помечен[2]. Использование помеченных и пустых квадратов означает, что мы допускаем разбиение нашего «окружения» (т. е. ленты) на части и возможность его описания множеством *дискретных* элементов (в противоположность непрерывному описанию). Это представляется вполне разумным, если мы хотим, чтобы наше устройство работало надежно и совершенно определенным образом. В силу используемой математической идеализации мы допускаем (потенциальную) бесконечность «окружения», однако в каждом *конкретном* случае входные данные, промежуточные вычисления и окончательный результат всегда должны быть *конечными*. Таким образом, хотя лента и имеет бесконечную длину, на ней должно быть конечное число непустых квадратов. Другими словами, и с той, и с другой стороны от устройства найдутся квадратики, после которых лента будет абсолютно пустой. Мы обозначим пустые квадраты символом «**0**», а помеченные — символом «**1**», например:

…	0	0	0	1	1	1	1	0	1	0	0	1	1	1	0	0	1	0	0	1	0	1	1	0	1	0	0	…

Нам нужно, чтобы устройство «считывало» информацию с ленты. Мы будем считать, что оно считывает по *одному* квадрату за раз и смещается после этого ровно на *один* квадрат влево или вправо. При этом мы не утрачиваем общности рассуждений: устройство, которое читает за один раз n квадратов или перемещается на k квадратов, легко моделируется устройством, указанным выше. Передвижение на k квадратов можно построить из k перемещений по одному квадрату, а считывание n квадратов за один прием сводится к запоминанию результатов n однократных считываний.

Что именно может делать такое устройство? Каким образом в самом общем случае могло бы функционировать устройство, названное нами «механическим»? Вспомним, что число *внутренних состояний* нашего устройства должно быть конечным. Все, что нам надо иметь в виду помимо этого — это то, что поведение нашего устройства полностью определяется его внутренним состоянием и входными данными. Входные данные мы упростили до двух символов — «**0**» и «**1**». При заданном начальном состоянии и таких входных данных устройство должно работать совершенно определенным образом: оно переходит в новое состояние (или остается в прежнем), заменяет считанный символ **0** или **1** тем же или другим символом **1** или **0**, передвигается на один квадрат вправо или влево, и наконец, оно решает, продолжить вычисления или же закончить их и остановиться.

---

[2] На самом деле Тьюринг использовал более сложные записи на ленте, но это не имеет принципиального значения, поскольку такие записи всегда могут быть представлены в виде последовательностей меток (одного типа) и пробелов. Я и далее, когда это допустимо, буду довольно свободно обращаться с исходными определениями Тьюринга.

Чтобы явно определить операции, производимые нашим устройством, для начала пронумеруем его внутренние состояния, например: 0, 1, 2, 3, 4, 5. Тогда действия нашего устройства, или машины Тьюринга, полностью определялись бы неким явным списком замен, например:

00 → 00R
01 → 131L
10 → 651R
11 → 10R
20 → 01R.STOP
21 → 661L
30 → 370R
. .
. .
. .
2100 → 31L
. .
. .
. .
2581 → 00R.STOP
2590 → 971R
2591 → 00R.STOP

Выделенная цифра слева от стрелки — это символ на ленте, который устройство в данный момент считывает. Оно заменяет этот символ выделенной цифрой в середине справа от стрелки. R означает, что устройство должно переместиться вдоль ленты на один квадрат *вправо*, а L соответствует такому же перемещению *влево*. (Если, в соответствии с исходным представлением Тьюринга, мы полагаем, что движется не устройство, а лента, то R означает перемещение *ленты* на один квадрат *влево*, а L — *вправо*.) Слово STOP означает, что вычисления завершены и устройство должно остановиться. Например, вторая инструкция 01 → 131L говорит о том, что если устройство находится в начальном состоянии 0 и считывает с ленты **1**, то оно должно перейти в состояние 13, оставить на ленте тот же символ **1** и переместиться по ленте на один квадрат влево. Последняя же инструкция 2591 → 00R.STOP говорит о том, что если устройство находится в состоянии 259 и считывает с ленты **1**, то оно должно вернуться в состояние 0, стереть с ленты **1**, т. е. записать в текущий квадрат **0**, переместиться по ленте на один квадрат вправо и прекратить вычисления.

Вместо номеров 0, 1, 2, 3, 4, 5, ... для обозначения внутренних состояний мы можем — и это более соответствовало бы знаковой системе нанесения меток на ленту — прибегнуть к системе нумерации, построенной только на символах «**0**» и «**1**». Состояние n можно было бы обозначить просто последовательностью из n единиц, но такая запись неэффективна. Вместо этого мы используем двоичную систему счисления, ставшую теперь общепринятой:

0 → 0,
1 → 1,
2 → 10,
3 → 11,
4 → 100,
5 → 101,
6 → 110,
7 → 111,
8 → 1000,
9 → 1001,
10 → 1010,
11 → 1011,
12 → 1100 и т. д.

Здесь последняя цифра справа соответствует «единицам» точно так же, как и в стандартной (десятичной) системе записи, но цифра прямо перед ней показывает число «двоек», а не «десятков». В свою очередь третья цифра справа относится не к «сотням», а к «четверкам»; четвертая — к «восьмеркам», а не к «тысячам» и т. д. При этом разрядность каждой последующей цифры (по мере продвижения влево) дается соответственной степенью двойки: 1, 2, 4 $(= 2 \times 2)$, 8 $(= 2 \times 2 \times 2)$, 16 $(= 2 \times 2 \times 2 \times 2)$, 32 $(= 2 \times 2 \times 2 \times 2 \times 2)$. (В дальнейшем нам будет иногда удобно использовать в качестве основания системы счисления числа, отличные от «2» и «10». Например, запись десятичного числа 64 по основанию «*три*» даст 2101, где каждая цифра теперь — некоторая степень тройки: $64 = (2 \times 3^3) + 3^2 + 1$; см. главу 4, сноску на с. 116.)

Используя двоичную запись для внутренних состояний, можно представить вышеприведенную инструкцию, описываю-

щую машину Тьюринга, следующим образом:

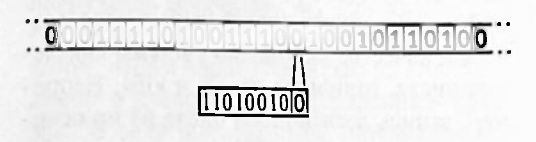
00 → 00R
01 → 11011L
10 → 1000001**1**R
11 → 10R
100 → 01STOP
101 → 10000101L
110 → 100101**0**R
. .
. .
. .
110100100 → 111L
. .
. .
. .
1000000101 → 00STOP
1000000110 → 1100001**1**R
1000000111 → 00STOP

Здесь я к тому же сократил R.STOP до STOP, поскольку мы вправе считать, что L.STOP никогда не происходит, так как результат последнего шага вычислений, будучи частью окончательного ответа, всегда отображается слева от устройства.

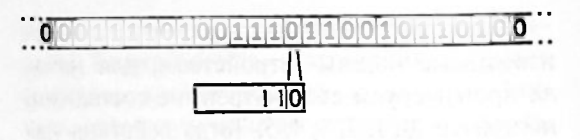
Предположим, что наше устройство находится во внутреннем состоянии, представленном бинарной последовательностью 11010010, и процессу вычисления соответствует участок ленты, изображенный на с. 66. Пусть мы задаем команду

110100100 → 11**1**L

Та цифра на ленте, которая в данный момент считывается (в нашем случае цифра «**0**»), показана «жирным» символом справа от последовательности нулей и единиц, обозначающих внутреннее состояние.



В частично описанном выше примере машины Тьюринга (который я выбрал более-менее произвольно) считанный «0» был бы тогда замещен на «**1**», внутреннее состояние поменялось бы на «11» и устройство переместилось бы на один шаг влево:



Теперь устройство готово к считыванию следующей цифры, снова «**0**». Согласно таблице, оно оставляет этот «**0**» нетронутым, но изменяет свое внутреннее состояние на «100101» и передвигается по ленте назад, т. е. на один шаг вправо. Теперь оно считывает «**1**» и находит где-то ниже в таблице инструкцию, которая определяет изменение внутреннего состояния и указывает, должна ли быть изменена считанная цифра и в каком направлении по ленте должно дальше двигаться устройство. Таким образом устройство будет действовать до тех пор, пока не достигнет команды STOP. В этой точке — после еще одного шага вправо — раздастся звонок, оповещающий оператора о том, что вычисления завершены.

Мы будем считать, что машина всегда начинает с внутреннего состояния «0» и что вся лента справа от устройства изначально пуста. Все инструкции и данные подаются в устройство с правой стороны. Как упоминалось ранее, эта информация всегда имеет форму конечной строки из нулей и единиц, за которой следует пустая лента (т. е. нули). Когда машина получает команду STOP, результаты вычислений оказываются на ленте слева от считывающего устройства.

Поскольку мы хотели бы иметь возможность вводить в устройство и числовые данные, то нам потребуется некий способ описания обычных чисел (под которыми я здесь имею в виду целые неотрицательные числа 0, 1, 2, 3, 4, ...) как части входной информации. Для представления числа n можно было бы просто использовать строку из n единиц (хотя при этом могут возникнуть трудности, когда речь зайдет о нуле):

1 → **1**,
2 → **11**,
3 → **111**,
4 → **1111**,
5 → **11111** и т. д.

Эта примитивная схема нумерации называется (хотя и довольно нелогично) *унарной* (*единичной*) системой. В этом случае

символ **0** мог бы использоваться в качестве пробела для разделения двух разных чисел. Наличие такого способа разделения для нас существенно, так как многие алгоритмы оперируют не отдельными числами, а *множествами* чисел. Например, для выполнения алгоритма Евклида наше устройство должно производить определенные действия над *парой* чисел A и B. Соответствующая машина Тьюринга может быть легко записана в явном виде. В качестве упражнения заинтересованный читатель может проверить, что нижеследующий набор инструкций действительно описывает машину Тьюринга (которую я буду называть **EUC**), выполняющую алгоритм Евклида, если в качестве исходных данных использовать два «унарных» числа, разделенных символом **0**:

00 → 00R
01 → 11L
10 → 101R
11 → 11L
100 → 10100R
101 → 110R
110 → 1000R
111 → 111R
1000 → 1000R
1001 → 1010R
1010 → 1110L
1011 → 1101L
1100 → 1100L
1101 → 11L
1110 → 1110L
1111 → 10001L
10000 → 10010L
10001 → 10001L
10010 → 100R
10011 → 11L
10100 → 00STOP
10101 → 10101R

Однако я бы порекомендовал такому читателю начать не с этого упражнения, а с чего-нибудь гораздо более простого, например, с машины Тьюринга **UN + 1**, которая просто прибавляет единицу к числу в унарном представлении:

00 → 00R
01 → 11R
10 → 01STOP
11 → 11R

Чтобы убедиться в том, что **UN + 1** на самом деле производит такую операцию, давайте мысленно применим ее, скажем, к ленте вида

...**00000111100000**...,

соответствующей числу четыре. Мы будем полагать, что наше устройство сначала находится где-то слева от последовательности единиц. Находясь в исходном состоянии 0, оно считывает **0**, в соответствии с первой инструкцией сохраняет его неизмененным, после чего перемещается на шаг вправо, оставаясь во внутреннем состоянии 0. Оно продолжает последовательно передвигаться вправо до тех пор, пока не встретит первую единицу. После этого вступает в силу вторая инструкция: устройство оставляет единицу как есть и сдвигается на шаг вправо, но уже в состоянии 1. В соответствии с четвертой инструкцией оно сохраняет внутреннее состояние 1, равно как и все считываемые единицы, двигаясь вправо до встречи с первым после набора единиц нулем. Тогда начинает действовать третья инструкция, согласно которой устройство заменяет этот нуль на **1**, перемещается на один шаг вправо (вспомним, что команда STOP эквивалентна R.STOP) и останавливается. Тем самым к последовательности из четырех единиц прибавляется еще одна, превращая — как и требовалось — 4 в 5.

В качестве несколько более трудного упражнения можно проверить, что машина **UN × 2**, определяемая набором инструкций

00 → 00R
01 → 10R
10 → 101L
11 → 11R
100 → 110R
101 → 1000R
110 → 01STOP
111 → 111R
1000 → 1011L
1001 → 1001R
1010 → 101L
1011 → 1011L

удваивает унарное число, как и должно быть, судя по ее названию.

Чтобы понять, как работает машина **EUC**, нужно явным образом задать пару подходящих чисел, скажем, 6 и 8. Как и ранее,

изначально машина находится во внутреннем состоянии 0 и расположена слева, а лента выглядит следующим образом:

...0000011111011111111100000....

После того, как машина Тьюринга после большого числа шагов останавливается, мы получаем ленту с записью вида

...00001100000000000...,

при этом машина располагается справа от ненулевых цифр. Таким образом, найденный наибольший общий делитель равен 2 (как и должно быть).

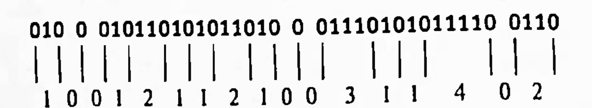
Исчерпывающее объяснение, *почему* машина **EUC** (или **UN** × **2**) на самом деле осуществляет действие, для которого она предназначена, включает в себя некоторые тонкости, и разобраться в нем, может быть, даже труднее, чем понять устройство самой машины — довольно обычная ситуация с компьютерными программами! (Чтобы полностью понять, почему алгоритмические процедуры делают то, что от них ожидается, необходима определенная интуиция. А не являются ли интуитивные прозрения сами алгоритмическими? Это один из вопросов, которые будут для нас важны в дальнейшем.) Я не буду пытаться дать здесь такое объяснение для приведенных примеров **EUC** или **UN** × **2**. Читатель, шаг за шагом проверив их действие, обнаружит, что я незначительно изменил обычный алгоритм Евклида, чтобы получить более компактную запись в рамках используемой схемы. И все же описание **EUC** остается достаточно сложным, включая в себя 22 элементарные инструкции для 11 различных внутренних состояний. В основном эти сложности носят чисто организационный характер. Можно отметить, например, что из этих 22 инструкций только 3 в действительности изменяют запись на ленте! (Даже для **UN** × **2** я использовал 12 инструкций, половина из которых меняют запись на ленте.)

Двоичная запись цифровых данных

Унарная система чрезвычайно неэффективна для записи больших чисел. Поэтому мы по большей части будем использовать вышеописанную *двоичную* систему. Однако, сделать это напрямую и попытаться читать ленту просто как двоичное число мы не сможем. Дело в том, что мы не имеем возможности сказать, когда кончается двоичное представление числа и начинается бесконечная последовательность нулей справа, которая отвечает пустой ленте. Нам нужен способ как-то обозначать конец двоичной записи числа. Более того, часто нам будет нужно вводить в машину *несколько* чисел, как, например, в случае с алгоритмом Евклида, когда требуется *пара* чисел [2]. Но в двоичном представлении мы не можем отличить *пробелы* между числами от нулей или строчек нулей, входящих в записи этих двоичных чисел. К тому же, помимо чисел нам может понадобиться и запись всевозможных сложных инструкций на той же ленте. Для того чтобы преодолеть эти трудности, воспользуемся процедурой, которую я буду в дальнейшем называть *сокращением* и согласно которой любая строчка нулей и единиц (с конечным числом единиц) не просто считывается как двоичное число, но замещается строкой из нулей, единиц, двоек, троек и т. д. таким образом, чтобы каждое число в получившейся строчке соответствовало числу единиц между соседними нулями в исходной записи двоичного числа. Например, последовательность

01000101101010110100011101010111100110

превратится в



Мы теперь можем считывать числа 2, 3, 4, ... как метки или инструкции определенного рода. Действительно, пусть 2 будет просто «запятой», указывающей на пробел между двумя числами, а числа 3, 4, 5, ... могли бы по нашему желанию символизировать различные инструкции или необходимые обозначения, как, например, «минус», «плюс», «умножить», «перейти в позицию со следующим числом», «повторить предыдущую операцию следующее число раз»,

и т. п. Теперь у нас есть разнообразные последовательности нулей и единиц, разделенные цифрами большей величины. Эти последовательности нулей и единиц будут представлять собой обычные числа, записанные в двоичной форме. Тогда записанная выше строка (при замене двоек «запятыми») примет вид:

(двоичное число **1001**) запятая (двоичное число **11**) запятая … .

Используя обычные арабские числа «9», «3», «4», «0» для записи соответствующих двоичных чисел **1001**, **11**, **100** и **0**, получаем новую запись всей последовательности в виде:

9, 3, 4 (инструкция 3) 3 (инструкция 4) 0.

Такая процедура дает нам, в частности, возможность указывать, где заканчивается запись числа (и тем самым отделять ее от бесконечной полосы пустой ленты справа), просто используя запятую в конце этой записи. Более того, она позволяет закодировать любую последовательность натуральных чисел, записанных в двоичной системе, как простую последовательность нулей и единиц, в которой для разделения чисел мы используем запятые. Посмотрим, как это сделать, на конкретном примере. Возьмем последовательность

5, 13, 0, 1, 1, 4.

В двоичном представлении она эквивалентна последовательности

101, 1101, 0, 1, 1, 100,

что на ленте можно записать с помощью операции *расширения* (обратной по отношению к описанной выше процедуре сокращения) как

…**0000100101101010010110011010110 10110100011000**… .

Такое кодирование легко выполнить, если в исходной двоичной записи чисел провести следующие замены:

0 → **0**
1 → **10**
, → **110**

и после этого добавить бесконечные последовательности нулей с обеих сторон вновь полученной записи. Чтобы сделать более понятной эту процедуру в применении к нашему примеру, разделим полученные двоичные числа пробелами:

0000 10 0 10 110 10 10 0 10 110 0 110 10 110 10 110 10 0 0 110 00.

Я буду называть этот способ представления (наборов) чисел *расширенной двоичной* записью. (Так, в частности, в расширенной двоичной форме записи число 13 выглядит как **1010010**.)

Есть еще одно, последнее, замечание, которое надо сделать в связи с этой системой записи. Это не более, чем техническая деталь, но она необходима для полноты изложения [3]. Двоичная (или десятичная) запись натуральных чисел в некоторой степени избыточна в том смысле, что нули, расположенные слева от записи числа, «не считаются» и обычно опускаются, так что **00110010** представляет собой то же самое двоичное число, что и **110010** (а 0050 — то же самое десятичное число, что и 50). Эта избыточность распространяется и на нуль, который может быть записан и как 000, и как 00, и, конечно, как 0. На самом деле и пустое поле, если рассуждать логически, должно обозначать нуль! В обычном представлении это привело бы к большой путанице, но в описанной выше системе кодирования никаких затруднений не возникает: нуль между двумя запятыми можно записать просто в виде двух запятых, следующих подряд (,,). На ленте такой записи будет соответствовать код, состоящий из двух пар единиц, разделенных одним нулем:

…**001101100**… .

Тогда исходный набор из шести чисел может быть записан в двоичной форме как

101, 1101,,1, 1, 100,

и на ленте при кодировании в расширенной двоичной форме мы получим последовательность

…**00001001011010100101101101011010110100011000**…,

в которой на один нуль меньше по сравнению с предыдущим кодом того же набора.

Теперь мы можем рассмотреть машину Тьюринга, реализующую, скажем, алгоритм Евклида в применении к паре чисел, записанных в расширенной бинарной форме. Для примера возьмем ту же пару чисел — 6 и 8, которую мы брали ранее. Вместо прежней унарной записи

$$\ldots 00000111111011111111100000 \ldots$$

воспользуемся двоичным представлением 6 и 8, т. е. 110 и 1000, соответственно. Тогда эта *пара* имеет вид

6, 8, или в двоичной форме **110, 1000,**

и в расширенной двоичной записи на ленте она будет выглядеть следующим образом

$$\ldots 00000101001101000011000000 \ldots .$$

Для этой конкретной пары чисел двоичная форма записи не дает никакого выигрыша по сравнению с унарной. Предположим, однако, что мы берем для вычислений (десятичные) числа 1 583 169 и 8610. В двоичной записи они имеют вид

110000010100001000001,
10000110100010.

На ленте при расширенном двоичном кодировании им будет соответствовать последовательность

$$\ldots 0010100000001001000001000000010110100000101001000010011 0 \ldots ,$$

которая занимает менее двух строк, тогда как для унарной записи пары чисел «1 583 169, 8610» не хватило бы места на страницах этой книги!

Машину Тьюринга, выполняющую алгоритм Евклида для чисел, записанных в расширенной двоичной форме, при желании можно получить из EUC с помощью пары дополнительных алгоритмов, которые переводили бы числа из расширенной двоичной формы в унарную и обратно. Однако, такой подход чрезвычайно неэффективен, ибо громоздкость унарной системы записи была бы по-прежнему «внутренне» присуща всему устройству, что проявилось бы в его низком быстродействии и потребности в огромном количестве «черновиков» (на левой стороне ленты). Можно построить и более эффективную машину Тьюринга для алгоритма Евклида, оперирующую исключительно расширенными двоичными числами, но для понимания принципов ее работы это не особенно важно.

Для того чтобы показать, каким образом машина Тьюринга может работать с числами в расширенном двоичном представлении, обратимся к значительно более простой, чем алгоритм Евклида, процедуре — просто *прибавлению единицы* к произвольному натуральному числу. Ее можно выполнить с помощью следующей машины Тьюринга (которую я назову **XN + 1**):

00 → 00R
01 → 11R
10 → 00R
11 → 101R
100 → 110L
101 → 101R
110 → 01STOP
111 → 1000L
1000 → 1011L
1001 → 1001L
1010 → 1100R
1011 → 101R
1101 → 1111R
1110 → 111R
1111 → 1110R

И вновь некоторые дотошные читатели могут захотеть проверить, вправду ли эта машина Тьюринга действует так, как должна, если взять, скажем, число 167. Это число имеет двоичное представление 10100111 и записывается на ленте как

$$\ldots 0000100100010101011000 \ldots .$$

Чтобы прибавить единицу к двоичному числу, мы просто находим в его записи последний нуль и меняем его на единицу, а все непосредственно следующие за ним единицы — на нули. Так что 167 + 1 = 168 в двоичной форме записывается в виде

$$10100111 + 1 = 10101000.$$

Таким образом, наша «прибавляющая единицу» машина Тьюринга должна превратить предыдущую запись на ленте в

$$\ldots 0000100100100001100000 \ldots ,$$

что она и делает.

Обратите внимание, что даже самая простая операция прибавления единицы в

такой записи выглядит довольно сложно, включая в себя 15 инструкций и восемь различных внутренних состояний! Конечно, в случае унарной записи все было значительно проще, поскольку тогда «прибавление единицы» означало удлинение строчки единиц еще на одну, поэтому не удивительно, что машина **UN + 1** была более простой. Однако, для очень больших чисел **UN + 1** была бы слишком медленной из-за чрезмерной длины ленты, и тогда более сложная машина **XN + 1**, но работающая с более компактным расширенным двоичным представлением, оказалась бы предпочтительнее.

Несколько отступая в сторону, я укажу операцию, для которой машина Тьюринга проще в расширенной двоичной, нежели в унарной форме — это *умножение на два*. Действительно, машина Тьюринга **XN × 2**, заданная в виде

00 → 00R
01 → 10R
10 → 01R
11 → 100R
100 → 111R
110 → 01STOP

запросто выполнит эту операцию в расширенной двоичной форме, тогда как соответствующая унарная машина **UN × 2**, описанная ранее, гораздо сложнее!

Этот раздел дает определенное представление о том, на что способны в простейших случаях машины Тьюринга. Как и следовало ожидать, при выполнении более или менее сложных операций эти машины могут становиться, и действительно становятся, несравненно более сложными. Каковы же принципиальные возможности таких устройств? Мы рассмотрим этот вопрос в следующем параграфе.

Тезис Черча—Тьюринга

После ознакомления с принципами построения простых машин Тьюринга легко убедиться, что все основные математические операции, такие как сложение двух чисел, их перемножение или возведение одного из них в степень другого, могут на самом деле быть выполнены соответствующими машинами Тьюринга. Построение таких машин в явном виде не представляет больших затруднений, но я не собираюсь сейчас этим заниматься. Машины Тьюринга могут выполнять операции, результат которых выражается парой натуральных чисел, например, деление с остатком, или сколь угодно большим, но конечным множеством чисел. Более того, можно сконструировать такие машины Тьюринга, для которых арифметические операции не предопределены заранее, а могут задаваться инструкциями, вводимыми с ленты. При этом возможно, что та конкретная операция, которая должна быть выполнена, будет зависеть в тот или иной момент от результатов вычислений, которые машина должна была выполнить на предыдущих этапах. («Если результат вычислений больше, чем то-то, надо сделать то-то, в противном случае выполнить то-то».) Убедившись, что можно построить машины Тьюринга, выполняющие арифметические или простые логические операции, уже не так трудно представить себе, какими должны быть машины, выполняющие более сложные задачи алгоритмического характера. «Повозившись» немного с подобными задачами, легко приходишь к убеждению в том, что машина этого типа может выполнять вообще *любые механические операции*! Тогда с точки зрения математики приобретает смысл *определение* механической операции как такой операции, которую может выполнить подобная машина. Существительное «алгоритм» и прилагательные «вычислимый», «рекурсивный» и «эффективный» используются математиками для обозначения механических операций, которые могут быть выполнены теоретическими устройствами такого рода, т. е. машинами Тьюринга. Если некоторая процедура четко определена и по природе своей механистична, то можно вполне обоснованно предположить, что найдется машина Тьюринга, способная ее выполнить. Это, в конце концов, и есть основной момент наших (то есть Тьюринга) рассуждений, лежащий и в основе самой концепции машины Тьюринга.

С другой стороны, остается ощущение, что принципы построения этих машин содержат излишние ограничения. Разрешение

устройству считывать за один раз только одну двоичную цифру (0 или 1) и передвигаться каждый раз только на один шаг да еще вдоль *единственной* одномерной ленты, на первый взгляд, ограничивает возможности машины. Почему бы не разрешить одновременное использование четырех, пяти или, возможно, тысячи разных лент, по которым одновременно двигалось бы большое количество взаимосвязанных считывающих устройств? Почему бы не ввести целую плоскость с нулями и единицами (или, например, трехмерное пространство), вместо того чтобы настаивать на использовании одномерной ленты? Почему бы не использовать другие системы счисления или символы из каких-нибудь более сложных алфавитов? По сути, ни одно из этих изменений ни в малейшей степени не влияет на то, что в принципе может быть достигнуто с помощью машины Тьюринга, хотя некоторые из них отразились бы на экономичности производимых операций (как это наверняка произошло бы, разреши мы использование нескольких лент). Класс осуществляемых операций, попадающих, таким образом, под определение «алгоритма» (или «вычисления», или «выполнимой процедуры», или «рекурсивной операции»), остался бы в точности тем же самым, если мы расширим определение наших машин и включим в него даже все предлагавшиеся выше модификации одновременно!

Мы можем видеть, что нет *необходимости* в дополнительных лентах, коль скоро устройство может по мере надобности находить свободное место на одной ленте. При этом может потребоваться постоянная перезапись данных с одного места ленты на другое. Это, может быть, «неэффективно», но в принципе не ограничивает возможности машин Тьюринга [4]. Сходным образом, использование более чем одного устройства Тьюринга для *параллельных вычислений* — идея, ставшая очень популярной в последние годы в связи с попытками более точного моделирования человеческого мозга, — не дает никаких принципиальных преимуществ (хотя при определенных обстоятельствах может увеличиться быстродействие). Использование двух непосредственно не связанных друг с другом устройств не даст выигрыша по сравнению с двумя взаимосвязанными устройствами. Но если два устройства связаны друг с другом, то, в сущности, это уже одно устройство!

А что можно сказать об ограничении Тьюринга, касающегося одномерности ленты? Если мы считаем, что эта лента представляет собой «окружение», то, возможно, мы бы предпочли в качестве такового иметь плоскую поверхность, или, допустим, трехмерное пространство. Может показаться, что плоскость лучше подошла бы для изображения «блок-схемы» вычислений (как в вышеприведенном описании последовательности действий алгоритма Евклида), чем одномерная лента [3]. Однако запись блок-схемы в «одномерной» форме не представляет принципиальных трудностей (например, можно использовать обычное словесное описание). Двумерное плоское изображение дает только удобство и простоту восприятия, но, по сути, ничего не меняет. Всегда есть возможность преобразовать координаты отметки или объекта на двумерной плоскости или в трехмерном пространстве и явным образом отобразить их на одномерной ленте. (Фактически, использование двумерной плоскости полностью эквивалентно использованию *двух* лент. Две ленты дают две «координаты», которые нужны для определения местоположения точки на двумерной плоскости; аналогично, *три* ленты могут выполнять ту же роль для точки в трехмерном пространстве.) И хотя эта одномерная запись может вновь оказаться «неэффективной», принципиальные возможности устройства это никак не ограничивает.

Несмотря на все это, по-прежнему остается вопрос о том, действительно ли понятие машины Тьюринга охватывает *все* логические или математические операции, которые мы могли бы назвать «механическими». В то

[3] В согласии с предложенным здесь описанием, эта блок-схема была бы скорее частью «устройства», нежели внешнего окружения — «ленты». На ленте мы до сих пор отображали только числа A, B, A − B, и т. п. Однако в дальнейшем нам потребуется также возможность описания и самого *устройства* в линейной одномерной форме. Как мы увидим далее в связи с *универсальной* машиной Тьюринга, есть тесная взаимосвязь между свойствами конкретного «устройства» и свойствами возможных «данных» (или «программы») для него. Поэтому удобно в *обоих* случаях придерживаться одномерной формы записи.

время, когда Тьюринг написал свою основополагающую работу, ситуация была гораздо менее ясной, чем сегодня, поэтому Тьюринг справедливо посчитал необходимым предоставить развернутое изложение этого вопроса. Детально рассмотренная Тьюрингом проблема получила дополнительное обоснование благодаря тому, что совершенно независимо от Тьюринга (и на самом деле несколько ранее) американский логик Алонзо Черч (совместно со Стивеном Клини), стремясь найти решение проблемы алгоритмической разрешимости Гильберта, предложил свою схему лямбда-исчисления. Хотя то, что это была всеобъемлющая полностью механическая схема, было не так очевидно, как в случае с подходом Тьюринга, ее несомненным преимуществом была удивительная компактность математической структуры. (Я буду рассматривать замечательный анализ Черча в конце главы.) Независимо от Тьюринга были предложены и другие подходы к решению задачи Гильберта (см. Ганди [1988]), среди которых можно выделить работу американского логика польского происхождения Эмиля Поста (опубликованную несколько позже работы Тьюринга, но содержащую идеи, более близкие идеям Тьюринга, нежели Черча). В скором времени было доказано, что все эти схемы совершенно эквивалентны.

Это значительно укрепило точку зрения, известную как *тезис Черча—Тьюринга*, которая утверждает, что машина Тьюринга (или ее эквивалент) на самом деле определяет то, что в математике понимают под алгоритмической (или выполнимой, или рекурсивной, или механической) процедурой. Сегодня, когда быстродействующие электронные компьютеры прочно вошли в нашу жизнь, немного найдется тех, кто считает необходимым ставить под сомнение эту теорию в ее изначальной формулировке. Вместо этого сейчас исследователи обратили внимание на вопрос, какие логические и математические операции могут выполнять реальные физические системы (возможно, включающие и человеческий мозг), подчиняющиеся точным физическим законам: точно такие же, что и машины Тьюринга, или же их возможности больше или меньше? Что касается меня, то я с удовольствием принимаю исходную *математическую* интерпретацию тезиса Черча—Тьюринга. С другой стороны, вопрос о его отношении к поведению реальных физических систем заслуживает отдельного рассмотрения и будет занимать в дальнейшем центральное место в наших рассуждениях.

Числа, отличные от натуральных

В предыдущих параграфах мы рассматривали действия над *натуральными* числами и отметили тот замечательный факт, что машина Тьюринга может оперировать с натуральными числами произвольной величины, несмотря на то, что каждая машина имеет фиксированное и *конечное* число внутренних состояний. Однако часто возникает необходимость в операциях с более сложными числами, такими как отрицательные числа, обыкновенные дроби и бесконечные десятичные дроби. Первые две категории (т. е. числа вида $-597/26$) легко поддаются обработке машинами Тьюринга, причем и числители, и знаменатели могут быть сколь угодно большими. Все, что для этого нужно — какой-нибудь подходящий код для знаков «−» и «/», который можно легко выбрать при использовании расширенной двоичной записи (например, «3» = **1110** для знака «−», а «4» = **11110** — для знака «/»). Таким образом, отрицательные числа и обыкновенные дроби рассматриваются как конечные наборы натуральных чисел, и с точки зрения общих вопросов вычислимости ничего нового не дают.

То же можно сказать и о *конечных* десятичных выражениях с произвольным числом знаков после запятой, поскольку они представляют собой лишь частный случай обыкновенных дробей. Так, например, конечная десятичная аппроксимация иррационального числа π, заданная числом 3,14159265, есть просто дробь 314 159 265/100 000 000. Однако *бесконечные* десятичные выражения, такие как полная запись числа π

$$\pi = 3{,}14159265358979\ldots,$$

представляют определенные трудности. На самом деле, ни входные, ни выходные дан-

ные машины Тьюринга не могут быть бесконечными десятичными выражениями. Можно было бы думать, что нашлась бы машина Тьюринга, способная выдавать одну за другой *все* последовательные цифры — 3, 1, 4, 1, 5, 9, ... в десятичной записи числа π и переносить их на выходную ленту, а мы просто позволим этой машине работать бесконечно долго. Но это *запрещено* для машин Тьюринга. Мы должны дождаться остановки машины (сопровождаемой звонком колокольчика!), прежде чем сможем ознакомиться с результатом. До того момента, пока машина не выполнит команды STOP, выходные данные могут изменяться и поэтому не являются достоверными. С другой стороны, после полной остановки машины результат должен быть с необходимостью конечным.

Существует, однако, «законная» процедура для того, чтобы заставить машину Тьюринга последовательно воспроизводить цифры примерно так, как это предлагалось выше. Если мы хотим получить бесконечную десятичную запись, скажем, числа π, мы могли бы заставить машину Тьюринга сначала рассчитать его целую часть, 3, используя на входе 0, затем — первую цифру дробной части, 1, используя на входе 1, затем — вторую цифру дробной части, 4, используя на входе 2, потом — третью цифру, 1, используя 3 и т. д. Вообще говоря, машина Тьюринга для получения всех цифр десятичной записи числа π в *этом* смысле *действительно* существует, хотя реализовать ее в явном виде было бы затруднительно. Подобное же замечание относится и ко многим другим иррациональным числам, таким, например, как $\sqrt{2} = 1{,}414213562\ldots$. Однако оказывается — и мы увидим это в следующей главе, — что некоторые иррациональные числа принципиально не могут быть получены с помощью машины Тьюринга. Числа, которые *можно* получить таким образом, называются *вычислимыми* (Тьюринг [1937]), а остальные (в действительности абсолютное большинство!) — *невычислимыми*. Я еще вернусь к этой теме и затрону ряд смежных вопросов в последующих главах. К нам это имеет отношение в связи с вопросом о том, может ли реальный физический объект (например, человеческий мозг) быть адекватно описан в терминах вычислимых математических структур в соответствии с нашими физическими теориями.

Проблема вычислимости важна для математики в целом. Не следует думать, что она относится только к *числам* как таковым. Ведь машины Тьюринга могут непосредственно оперировать *математическими формулами*, например, алгебраическими или тригонометрическими выражениями, или выполнять формальные действия математического анализа. Все, что для этого нужно, это некий способ точного кодирования всех используемых математических символов в виде последовательностей нулей и единиц, которые позволят применить соответствующую машину Тьюринга. Именно это Тьюринг имел в виду, когда он взялся за *проблему алгоритмической разрешимости*, в которой требуется найти алгоритмическую процедуру для ответа на самые *общие* математические вопросы. Очень скоро мы вновь обратимся к этой теме.

Универсальная машина Тьюринга

Я еще не затрагивал понятия *универсальной* машины Тьюринга. Лежащий в ее основе принцип понять нетрудно, хотя детали могут быть сложны. Основная идея состоит в том, чтобы закодировать команды для произвольной машины Тьюринга T в виде последовательности нулей и единиц, которую можно записать на ленте. Эта запись используется как начальная часть входных данных для некоторой *особой* машины Тьюринга U, называемой универсальной, которая затем обрабатывает остальную часть ленты в точности так, как это сделала бы машина T. Универсальная машина Тьюринга — это универсальный имитатор. Начальная часть ленты дает универсальной машине U всю информацию, необходимую для точной имитации любой машины T!

Чтобы показать, как это может быть реализовано, нам потребуется какая-нибудь система *нумерации* машин Тьюринга. Рассмотрим список инструкций, определяющих

произвольную машину Тьюринга, например, одну из описанных выше. Мы должны в соответствии с некоторыми четкими правилами представить эти инструкции в виде последовательностей нулей и единиц. Это можно сделать, например, с помощью процедуры «сокращения», которую мы использовали ранее. Тогда, если мы закодируем символы R, L, STOP, «стрелка» (→) и «запятая», скажем, числами 2, 3, 4, 5 и 6 соответственно, то мы сможем записать их в виде «сокращений» **110**, **1110**, **11110**, **111110** и **1111110**. Цифры 0 и 1, кодируемые, соответственно, как **0** и **10**, могут быть использованы для записи строк этих символов, входящих в таблицу действий машины Тьюринга. Нам не нужны различные обозначения для «жирных» цифр **0** и **1** и для остальных цифр в таблице, поскольку расположение «жирных» цифр в конце двоичного кода является достаточным отличительным признаком. При этом 1101, например, будет читаться как двоичное число 1101, представляемое на ленте последовательностью **1010010**. В частности, 00 будет читаться как 00, что без всякой двусмысленности можно закодировать как **0** или вовсе опустить. Можно существенно сэкономить, если не кодировать «стрелки» и непосредственно предшествующие им символы, а воспользоваться цифровым упорядочением команд, позволяющим определить, какими должны быть эти символы. Правда, для этого надо убедиться в отсутствии «дырок» в получившемся порядке и добавить, где требуется, «немые» команды. (Например, машина Тьюринга XN + 1 не имеет команды, соответствующей коду 1100, поскольку такая комбинация в ходе ее работы никогда не встречается. Следовательно, мы должны ввести в список команд немую команду, скажем 1100 → 00R, которая не вызовет каких бы то ни было изменений в работе машины. Сходным образом мы должны добавить немую команду 101 → 00R в список команд машины XN × 2.) Без таких «немых» команд кодирование последующих команд было бы нарушено. Как можно видеть, на самом деле мы не нуждаемся и в запятой в конце каждой команды, поскольку символов L и R вполне достаточно для отделения команд друг от друга. Поэтому мы просто будем использовать такую систему кодирования:

0 для 0 или **0**,
10 для 1 или **1**,
110 для R,
1110 для L,
11110 для STOP

В качестве примера выпишем команды для машины Тьюринга XN + 1 (с дополнительной немой командой 1100 → 00R). Опуская стрелки, цифры, непосредственно предшествующие им, и запятые, получим

00R	11R	00R	101R
110L	101R	01STOP	1000L
1011L	1001L	1100R	101R
00R	1111R	111R	1110R

Мы можем улучшить полученный результат, если опустим все 00 и заменим каждые 01 просто единицей в соответствии с тем, что говорилось ранее. Тогда мы получим строку символов

R11RR101R110L101R1STOP1000L1011L1001L1100R101RR1111R111R1110R,

которая на ленте записывается как последовательность

11010101101101001011010100111010010110101111010000111010010101110100010111010100011010010110110101010101101010101101010100110.

Есть еще два способа немного сэкономить. Во-первых, всегда можно удалить код **110** в начале записи (вместе с бесконечным участком пустой ленты, предшествующим этому коду). Он обозначает последовательность 00R, соответствующую начальной команде 00 → 00R, которую я до сих пор неявно считал общей для *всех* машин Тьюринга, поскольку она необходима для того, чтобы устройство, начав работу в произвольной точке слева от начала записи на ленте, могло перемещаться вправо до тех пор, пока не встретит первую непустую клетку. Во-вторых, точно так же всегда можно удалить код **110** (и неявную бесконечную последовательность нулей, которая, по предположению, следует за ним) в конце записи,

поскольку этой кодовой последовательностью должно заканчиваться описание *любой* машины Тьюринга (во всех случаях список команд заканчивается командой R, L или STOP). Получающееся *двоичное число* — это *номер* машины Тьюринга, который для XN + 1 будет выглядеть так:

10101101101001011010100111010010110101111010000111010010101110100010111010100011010010110110101010101101010101101010100.

В обычной десятичной записи этот номер равен

450 813 704 461 563 958 982 113 775 643 437 908.

Иногда машину с номером n мы, не вполне точно, будем называть n-й машиной Тьюринга и обозначать ее T_n. В этом случае XN + 1 становится

450 813 704 461 563 958 982 113 775 643 437 908-й

машиной Тьюринга!

Кажется поразительным факт, что нам надо пробежать так долго вдоль «списка» машин Тьюринга, чтобы найти машину, выполняющую такую тривиальную операцию, как прибавление единицы к натуральному числу (в расширенном двоичном представлении). (Я не думаю, что моя система кодирования была в целом настолько неэффективна, хотя в ней и есть еще возможности для незначительных улучшений.) В действительности, есть машины Тьюринга и с меньшими номерами, которые представляют интерес, например UN + 1 с двоичным номером

101011010111101010,

который в десятичной записи превращается всего лишь в 177 642. Значит, особенно тривиальная машина UN + 1, которая просто дописывает **1** единицу в конце последовательности единиц, является 177 642-й машиной Тьюринга. Интересно, что «умножение на два» в списке машин Тьюринга попадает где-то между этими двумя машинами, причем и в унарном, и в расширенном двоичном представлении: номер XN × 2 равен 10 389 728 107, а номер UN × 2 — 1 492 923 420 919 872 026 917 547 669.

Наверное, принимая во внимание величины этих номеров, уже не вызовет удивления тот факт, что абсолютное большинство натуральных чисел не соответствует ни одной рабочей машине Тьюринга. Приведем перечень первых тринадцати машин Тьюринга в соответствии с принятой нумерацией:

T_0: 00 → 00R, 01 → 00R,
T_1: 00 → 00R, 01 → 00L,
T_2: 00 → 00R, 01 → 01R,
T_3: 00 → 00R, 01 → 00STOP,
T_4: 00 → 00R, 01 → 10R,
T_5: 00 → 00R, 01 → 01L,
T_6: 00 → 00R, 01 → 00R, 10 → 00R,
T_7: 00 → 00R, 01 → ???,
T_8: 00 → 00R, 01 → 100R,
T_9: 00 → 00R, 01 → 10L,
T_{10}: 00 → 00R, 01 → 11R,
T_{11}: 00 → 00R, 01 → 01STOP,
T_{12}: 00 → 00R, 01 → 00R, 10 → 00R.

Из этих машин T_0 просто перемещается вправо, стирая все, что ей попадается на пути, никогда не останавливаясь и не меняя направления движения. Машина T_1 выполняет в сущности ту же операцию, но более громоздким путем, отступая на шаг назад каждый раз, когда она стирает очередную единицу на ленте. Так же как и T_0, машина T_2 двигается вправо, никогда не останавливаясь, но относится к ленте более «почтительно», попросту оставляя всю информацию нетронутой. Эти машины не могут использоваться в качестве машин Тьюринга, поскольку никогда не останавливаются. T_3 — первая в этом списке «правильная» машина: она скромно прекращает действие после того, как изменяет первую (самую левую) единицу на нуль. T_4 сталкивается с серьезной проблемой. Найдя первую единицу на ленте, она переходит во внутреннее состояние, которое нигде не описано, и, следовательно, машина не имеет никаких команд для следующего шага. С той же проблемой сталкиваются T_8, T_9 и T_{10}. С T_7 возникают трудности еще более фундаментального характера. Строка нулей и единиц, которой она представляется, включает последовательность из *пяти* единиц: **110111110**. Интерпретации этой последовательности не существует, поэтому T_7 намертво застревает сразу же, как

только доходит до первой единицы. (Я буду называть T_7, равно как и любую другую машину T_n, двоичное расширенное представлений которой содержит более четырех единиц, *некорректно определенной*.) Машины T_5, T_6 и T_{12} испытывают те же трудности, что и T_0, T_1, T_2: они просто никогда не останавливаются. Все эти машины — T_0, T_1, T_2, T_5, T_6, T_7, T_8, T_9, T_{10} и T_{12} — совершенно бесполезные устройства! Только T_3 и T_{11} являются функциональными машинами Тьюринга, да и то не слишком интересными. Причем T_{11} даже скромнее, чем T_3: натолкнувшись на первую же единицу, она останавливается и вообще ничего не меняет!

Надо заметить, что наш перечень содержит избыточную информацию. Машина T_{12} идентична T_6, а по действиям обе они аналогичны T_0, поскольку ни T_6, ни T_{12} никогда не переходят во внутреннее состояние 1. Но нам нет нужды волноваться из-за этой избыточности, равно как из-за изобилия неработоспособных (фиктивных) машин Тьюринга в нашем списке. На самом деле, мы могли бы изменить систему кодирования таким образом, чтобы избавиться от большого числа бесполезных устройств и значительно уменьшить избыточность списка машин. Но все это можно сделать только ценой усложнения нашей примитивной универсальной машины Тьюринга, которая должна расшифровывать вводимую в нее запись и имитировать машину T_n, чей номер она считала. Это было бы оправдано, если бы было можно избавиться от *всех* бесполезных (и повторяющихся) машин. Но это, как мы увидим чуть позднее, *невозможно*! Поэтому мы оставим нашу систему кодирования без изменений.

Будет удобно интерпретировать ленту с последовательностью меток на ней, например

$$\ldots \mathbf{0001101110010000} \ldots,$$

как двоичное представление некоторого числа. Вспомним, что нули простираются бесконечно в обе стороны, а вот количество единиц конечно. Кроме того, я буду полагать, что их число *отлично от нуля* (т. е. что в этой последовательности существует хотя бы одна единица). Мы можем тогда считывать конечную строку символов между первой и последней единицами (включительно), которая в предыдущем случае имеет вид

$$\mathbf{110111001},$$

как двоичное представление натурального числа (в десятичной форме это 441). Однако такая процедура даст нам только *нечетные* числа (их двоичное представление оканчивается на **1**), тогда как нам нужна возможность представления *всех* натуральных чисел. Поэтому мы воспользуемся следующим несложным приемом — будем удалять последнюю единицу (которая принимается просто за маркер, обозначающий конец выражения) и считывать оставшуюся часть как двоичное число [5]. Тогда в последнем примере получим двоичное число

$$\mathbf{11011100},$$

которое соответствует десятичному числу 220. Эта процедура имеет то преимущество, что нуль также представляется непустой лентой, а именно:

$$\ldots \mathbf{0000001000000} \ldots.$$

Рассмотрим, как действует машина Тьюринга T_n на некоторую (конечную) строку нулей и единиц на ленте, которая подается в устройство справа. Удобно рассматривать эту строку как двоичное представление некоторого числа, например m, в соответствии с приведенной выше схемой. Предположим, что после определенного числа шагов машина T_n в конце концов останавливается (т. е. доходит до команды STOP). Строка двоичных цифр, которые машина выписала к этому моменту на левой части ленты, и будет искомым результатом вычислений. Считывая эту последовательность в соответствии с той же схемой так же как двоичное представление некоторого числа, получим новое число, скажем, p. Тогда мы можем записать соотношение, выражающее тот факт, что результатом действия n-й машины Тьюринга T_n на число m является число p, следующим образом:

$$T_n(m) = p.$$

Взглянем на это соотношение с несколько иной точки зрения. Мы будем считать, что это выражение описывает некоторую специфическую операцию, которая

применяется к *паре* чисел n и m для того, чтобы получить p. (Это означает: для заданных *двух* чисел m и n мы можем найти значение p, если введем m в n-ю машину Тьюринга.) Эта специфическая операция является полностью алгоритмической. Поэтому она может быть выполнена *одной конкретной* машиной Тьюринга U; иными словами, U, совершая действие над *парой* (n, m), дает в результате p. Поскольку машина U должна производить операцию над обоими числами n и m, чтобы получить ответ, выражаемый одним числом p, то нам нужно придумать способ для записи пары (n, m) на *одной* ленте. С этой целью предположим, что n записывается в стандартной двоичной форме и заканчивается последовательностью **111110**. (Вспомним, что двоичный номер всякой корректно определенной машины Тьюринга, — это последовательность символов, состоящая только из сочетаний вида **0**, **10**, **110**, **1110** и **11110**, поэтому он нигде не содержит более четырех единиц подряд. Таким образом, если T_n — корректно определенная машина, то появление последовательности **111110** действительно будет означать конец записи номера n.) Все, что следует за ней, должно быть просто записью числа m на ленте в соответствии с приведенными выше правилами (т. е. двоичное число m и строка **1000**... непосредственно за ним). Таким образом, с этой второй частью ленты машина T_n и должна производить предполагаемые действия.

Если в качестве примера мы возьмем $n = 11$ и $m = 6$, то на ленте, вводимой в машину U, мы будем иметь последовательность

...**000101111111011010000**... .

Она образована из следующих составляющих:

...**0000** (пустое начало ленты)

1011 (двоичное представление одиннадцати)

111110 (обозначает окончание числа n)

110 (двоичное представление шести)

10000... (остаток ленты)

То, что машина Тьюринга U должна была бы делать на каждом очередном шагу процедуры, выполняемой T_n над m — это исследовать структуру последовательности цифр в выражении n с тем, чтобы можно было произвести соответствующие изменения цифр числа m (т. е. «ленты» машины T_n). В принципе, реализация такой машины не вызывает существенных затруднений (хотя и довольно громоздка на практике). Список ее собственных команд должен был бы просто содержать правила для чтения подходящей команды из «списка», закодированного в числе n, на каждом этапе выполнения действий над цифрами, считанными с «ленты», как они фигурируют в числе m. Можно предположить, что при этом совершалось бы значительное количество прыжков взад-вперед по ленте между цифрами, составляющими n и m, и выполнение процедуры было бы чрезвычайно медленным. Тем не менее, список команд подобной машины, несомненно, можно составить, и такая машина называется нами *универсальной* машиной Тьюринга. Обозначая ее действие на пару чисел (n, m) через $U(n, m)$, мы получаем:

$$U(n, m) = T_n(m)$$

при любых (n, m), для которых T_n — корректно определенная машина Тьюринга [6]. Машина U, в которую первым вводится число n, в точности имитирует n-ю машину Тьюринга!

Поскольку U — машина Тьюринга, то она сама будет иметь номер. То есть, для некоторого числа u имеем

$$U = T_u.$$

Сколь велико u? В сущности, мы можем положить, что *u в точности* равно следующему числу:

$u =$ 7244855335339317577198395039615711237952360672556559631108144796606505059404241090310483613632359365644443458382226883278767626556144692814117715017842551707554085657689753346356942478488597046934725739988582283827795294683460521061169835945938791885546326440925525505820555989451890716537414896033096753020431553625034984529832320651583047664142130708819329717234151056980262734686429921838172157333482823073453713421475059740345184372359593090640024321077342178851492760797597634415123079586396354492269159479654 61

4711345700145048167337562172573464522731
0544829807849651269887889645697609066342
0447798902191443793283001949357096392170
3904833270882596201301773727202718625919
9144282754374223513556751340842222998893
7441053430547104436869587640517812801943
7530813870639942772823156425289237514565
4438990527807932411448261423572861931183
3261065612275553181020751108533763380603
1082361675045635852164214869542347187426
4375444287900624858270912404220765387542
6445413345174856629157429990950262300973
3738137724162172747723610206786854002893
5660856968226201419824862169890260913094
0298570600174300670086896759034473417412
7874255812015493663938996905817738591654
0553567040928213322216314109787108145997
8669599704509681841906299443656015145490
4880922084480034822492077304030431884298
9939313526688234966210194716191070146196
8523192847482034495897709553561107027581
7487333272966789987984732840981907648512
7263100174016678736347760585724503696443
4897992034489997455662402937487668839751
4044516657077500605138839916688140725455
4466522205072426239237921152531816251253
6305093172863142200406457130527580230766
5183351995689139748137504926429605010013
651980186945639498

(или какому-нибудь другому подходящему, не менее внушительному по величине числу). Это число, без сомнения, выглядит устрашающе большим! Оно, действительно, чрезвычайно велико, но я не вижу способа, как его можно было бы сделать меньше. Процедуры кодирования и определения, использованные мною для машин Тьюринга, вполне разумны и достаточно просты, и все же с неизбежностью приводят к подобным несуразно большим числам для реальной универсальной машины Тьюринга [7].

Я уже говорил, что все современные общеупотребительные компьютеры, по сути, являются универсальными машинами Тьюринга. Я ни в коем случае не подразумеваю под этим, что их логическая структура должна в точности походить на предложенную мной выше структуру универсальной машины Тьюринга. Однако суть дела состоит в том, что если сперва ввести в произвольную универсальную машину Тьюринга соответствующую программу (начало подаваемой на вход ленты), то потом она сможет копировать поведение любой машины Тьюринга! В предыдущем примере программа просто принимает форму одного числа (числа n), но этим разнообразие возможных процедур и вариантов исходной схемы Тьюринга отнюдь не исчерпывается. В действительности я сам, описывая машину, несколько отклонился от того, что исходно было предложено Тьюрингом. Но ни одно из этих отклонений не имеет сейчас для нас существенного значения.

Неразрешимость проблемы Гильберта

Мы теперь вплотную подходим к той цели, ради которой Тьюринг с самого начала разрабатывал свою теорию — получить ответ на вопрос, заключенный в общей проблеме алгоритмической разрешимости, поставленной Гильбертом, а именно: существует ли некая механическая процедура для решения всех математических задач, принадлежащих к некоторому широкому, но вполне определенному классу? Тьюринг обнаружил, что он мог бы перефразировать этот вопрос следующим образом: *остановится* ли в действительности n-я машина Тьюринга, если на ее вход поступит число m? Эта задача получила название *проблемы остановки*. Не так сложно составить список команд, для которых машина никогда не остановится при *любом* m (как, например, в случаях $n = 1$ или 2, рассмотренных в предыдущем разделе, а также во всех случаях, когда вообще отсутствует команда STOP). Точно так же существует множество списков команд, для которых машина будет останавливаться всегда, независимо от вводимого числа m (например, T_{11}). Кроме того, некоторые машины при работе с одними числами останавливались бы, а с другими — нет. Совершенно очевидно, что алгоритм, который никогда не прекращает работу, бесполезен. Это, собственно, и не алгоритм вовсе. Поэтому важно уметь ответить на вопрос, приведет ли когда-нибудь работа машины T_n над данным числом m к какому-то ответу или нет! Если *нет* (т. е. процесс вычисления никогда не прекращается), то я буду

выражать это следующей записью:

$$T_n(m) = \square.$$

(Сюда же включены машины, которые в ходе работы попадают в ситуацию, когда нет команды, определяющей их дальнейшее поведение, как это было в случае рассмотренных выше фиктивных машин T_4 и T_7. К сожалению, наша на первый взгляд работоспособная машина T_3 должна теперь также считаться фиктивной, т. е. $T_3(m) = \square$, поскольку результатом ее действия всегда будет просто пустая лента, тогда как нам, чтобы приписать номер полученному ответу, нужна хотя бы одна единица на выходе! Машина T_{11}, однако, совершенно полноправна, поскольку она производит единственную **1**. Результатом ее работы будет лента с номером 0, так что $T_{11}(m) = 0$ для любого m.)

В математике весьма важно иметь возможность установить момент, когда машина Тьюринга остановится. Рассмотрим для примера уравнение

$$(x+1)^{w+3} + (y+1)^{w+3} = (z+1)^{w+3}.$$

(Не пугайтесь, даже если Вы не любите вникать в детали математических вычислений. Это уравнение используется здесь только в качестве примера, и от вас не требуется его глубокого понимания.) Это конкретное уравнение относится к известной (возможно, самой известной) и пока нерешенной математической проблеме. Проблема формулируется следующим образом: существует ли какой-либо набор x, y, z, w, для которого это равенство выполняется. Знаменитое утверждение, записанное на полях «Арифметики» Диофанта великим французским математиком семнадцатого столетия Пьером де Ферма (1601–1665) и известное как «последняя теорема Ферма», гласит, что это равенство *никогда* не выполняется[4] [8]. Будучи адвокатом по профессии, Ферма тем не менее был искуснейшим математиком своего времени. (Ферма был современником Декарта.) В своей записи он утверждал, что знает «воистину прекрасное доказательство» своей теоремы, но поля книги слишком малы, чтобы его привести. До сегодняшнего дня никому так и не удалось ни воспроизвести это доказательство[5], ни найти опровергающий это утверждение пример!

Очевидно, что для *заданной* четверки чисел (x, y, z, w) выяснить, выполняется это равенство или нет, можно простым вычислением. Значит, мы можем представить себе вычислительный алгоритм, который последовательно перебирает все возможные четверки чисел одну за другой и останавливается только тогда, когда равенство удовлетворяется. (Мы уже знаем, что для конечных наборов чисел существуют способы их кодирования на ленте вычислимым способом, а именно, в виде одного числа. Таким образом, перебор всех четверок можно провести, просто следуя естественному порядку соответствующих им одиночных чисел.) Если бы мы могли установить, что этот алгоритм никогда *не* останавливается, то это стало бы доказательством утверждения Ферма.

Сходным образом в терминах проблемы остановки машины Тьюринга можно перефразировать многие другие нерешенные математические проблемы. Примером такого рода проблем может служить так называемое предположение Гольдбаха: любое четное число, большее двух, может быть представлено в виде суммы двух простых чисел[6]. Процесс, с помощью которого можно установить, относится некоторое натуральное число к простым или нет, является алгоритмическим, поскольку достаточно проверить делимость данного числа на все числа, *меньшие* его, а это достигается с помощью *конечного* числа вычислительных операций. Мы можем придумать машину Тьюринга, которая перебирает четные числа 6, 8, 10, 12, 14, ..., пробуя все возможные способы разбиения их на пары нечетных чисел

$$6 = 3+3, \quad 8 = 3+5, \quad 10 = 3+7 = 5+5,$$
$$12 = 5+7, \quad 14 = 3+11 = 7+7, \quad \ldots$$

[4] Напомним, что *натуральными* мы называем числа 0, 1, 2, 3, 4, 5, 6, ... Вместо обычной записи ($x^w + y^w = z^w$, где $x, y, z > 0$, $w > 2$) мы используем «$x+1$», «$w+3$» и т. д., чтобы включить в рассмотрение все натуральные числа, начиная с нуля.

[5] Последняя теорема Ферма доказана английским математиком Эндрю Уайлсом (Andrew J. Wiles). Доказательство опубликовано в 1995 году. — *Прим. ред.*

[6] Напомним, что *простые* числа 2, 3, 5, 7, 11, 13, 17, ... — это такие натуральные числа, которые делятся только на самих себя и на единицу. Ни нуль, ни единица простыми числами не считаются.

и убеждаясь, что для *каждого* четного числа *какое-то* из разбиений образовано *двумя* простыми числами. (Очевидно, нам не надо проверять пары четных слагаемых, кроме $2+2$, поскольку все простые числа за исключением 2 — нечетные.) Наша машина должна остановиться только в том случае, если она находит четное число, для которого *ни одно* из разбиений не является парой простых чисел. В этом случае мы получили бы контрпример к предположению Гольдбаха, т. е. нашли бы четное число, большее 2, которое не является суммой двух простых чисел. Следовательно, если бы мы могли установить, останавливается машина Тьюринга когда-нибудь или нет, то тем самым мы выяснили бы, справедливо предположение Гольдбаха или нет.

Возникает естественный вопрос: каким образом следует определять, остановится какая-то определенная машина Тьюринга (в которую введены конкретные начальные данные) или нет? Для многих машин Тьюринга ответить на этот вопрос нетрудно, но, как мы видели выше, иногда для ответа может потребоваться решение какой-нибудь до сих пор не решенной математической задачи. Так существует ли некая *алгоритмическая* процедура для решения общей проблемы — проблемы остановки — полностью механическим путем? Тьюринг показал, что такой процедуры на самом деле нет.

В сущности, его доказательство сводилось к следующему. Предположим, наоборот, что указанный алгоритм существует[7]. Тогда существует и некая машина Тьюринга H, которая «решает», остановится ли в конце концов n-я машина Тьюринга, действуя на число m. Условимся, что результатом действия машины H будет лента с номером **0**, если n-я машина не останавливается, и с номером **1** в противоположном случае:

$$H(n;m) = \begin{cases} 0, & \text{если } T_n(m) = \square, \\ 1, & \text{если } T_n(m) \\ & \text{останавливается.} \end{cases}$$

[7] Это хорошо известный и очень мощный метод математического доказательства, называемый «доказательством от противного» или *reductio ad absurdum* (сведение к абсурду), в котором сначала полагается истинным утверждение, исключающее исходное, затем из этой предпосылки выводится противоречие, которое и служит доказательством справедливости исходного утверждения.

Здесь мы могли бы воспользоваться способом кодирования пары (n, m), использованным ранее для универсальной машины Тьюринга U. Однако это привело бы к проблеме технического характера, поскольку при некоторых n (например, $n = 7$) T_n будет определена некорректно, и маркер **111101** будет непригоден для отделения на ленте n от m. Чтобы избежать этой проблемы, будем полагать, что n представлено не в двоичной, а в *расширенной* двоичной форме, тогда как для m будет по-прежнему использоваться обычная двоичная запись. В этом случае комбинации **110** будет достаточно для разделения n и m. Использование точки с запятой в обозначении $H(n;m)$ в отличие от запятой в обозначении универсальной машины $U(n, m)$ указывает на это различие в кодировании.

Представим себе теперь бесконечную таблицу, в которую включены окончательные результаты действий всех возможных машин Тьюринга на все возможные (различные) входные данные. В этой таблице N-й ряд представляет собой результаты вычислений n-й машины Тьюринга, полученные при ее работе последовательно с $m = 0, 1, 2, 3, 4, \ldots$:

$m \rightarrow$	0	1	2	3	4	5	6	7	8	…
$n \downarrow$										
0	□	□	□	□	□	□	□	□	□	…
1	0	0	0	0	0	0	0	0	0	…
2	1	1	1	1	1	1	1	1	1	…
3	0	2	0	2	0	2	0	2	0	…
4	1	1	1	1	1	1	1	1	1	…
5	0	□	0	□	0	□	0	□	0	…
6	0	□	1	□	2	□	3	□	4	…
7	0	1	2	3	4	5	6	7	8	…
8	□	1	□	□	1	□	□	□	1	…
.	.				.				.	…
.	.				.				.	
.	.				.				.	
197	2	3	5	7	11	13	17	19	23	…
.	.				.				.	
.	.				.				.	
.	.				.				.	

Я немного «сжульничал» и не стал располагать машины Тьюринга по порядку их *действительных* номеров. Если бы я так сделал, то получился бы список, начало кото-

рого выглядело бы слишком скучным, поскольку все машины при значениях n меньших 11 не дают ничего, кроме □, а для $n = 11$ мы имеем просто нули. Дабы сделать начало этой таблицы более интересным, я предположил, что мы использовали некую гораздо более эффективную систему кодирования. Фактически, я просто присвоил ячейкам более или менее произвольные значения, только чтобы дать вам общее представление о том, как может выглядеть эта таблица.

На самом деле нам не требуется, чтобы эта таблица была построена путем *вычислений*, скажем, с помощью некоторого алгоритма. (На самом деле, как мы увидим далее, такого алгоритма и не существует.) Достаточно просто *представить* себе, что каким-то образом *истинный* список попал в наше распоряжение, возможно, с помощью Бога! Если бы мы попытались получить эту таблицу с помощью вычислений, то именно символы □ вызвали бы затруднения, поскольку мы не могли бы с уверенностью сказать, когда в той или иной ячейке должен быть помещен символ □ — ведь соответствующие вычисления никогда не заканчиваются!

Тем не менее искомую таблицу *можно* построить с помощью вычислительной процедуры, если использовать нашу гипотетическую машину H, поскольку она могла бы определить, где на самом деле появляются значения □. Однако вместо этого мы используем машину H для того, чтобы *избавиться* от появления значений □ в таблице, заменив их во всех случаях нулями. Это достигается за счет вычисления значения $H(n;m)$, предваряющего действие T_n на m, после чего мы позволим T_n производить соответствующие действия, только если $H(n;m) = 1$ (т. е. только тогда, когда вычисление $T_n(m)$ приводит к определенному результату), и будем просто записывать в соответствующую ячейку 0 при $H(n;m) = 0$ (т. е. если $T_n(m) = □$). Мы можем записать эту новую процедуру, представляющую собой последовательное действие $H(n;m)$ и $T_n(m)$, как

$$T_n(m) \times H(n;m).$$

(Здесь я использую общепринятую в математике договоренность о последовательности выполнения действий, согласно которой операция, записанная *справа*, должна выполняться *первой*. Обратите внимание, что в этом случае можно символически записать □ × 0 = 0.)

Теперь таблица принимает следующий вид:

$n \downarrow$ \ $m \rightarrow$	0	1	2	3	4	5	6	7	8	...
0	0	0	0	0	0	0	0	0	0	...
1	0	0	0	0	0	0	0	0	0	...
2	1	1	1	1	1	1	1	1	1	...
3	0	2	0	2	0	2	0	2	0	...
4	1	1	1	1	1	1	1	1	1	...
5	0	0	0	0	0	0	0	0	0	...
6	0	0	1	0	2	0	3	0	4	...
7	0	1	2	3	4	5	6	7	8	...
8	0	1	0	0	1	0	0	0	1	...
⋮	⋮				⋮				⋮	

Заметьте, что, исходя из предположения существования машины H, мы получаем ряды таблицы, состоящие из *вычислимых* последовательностей. (Под «вычислимой последовательностью» я понимаю бесконечную последовательность, элементы могут быть найдены один за другим посредством некоего алгоритма; это означает, что существует некоторая машина Тьюринга, которая, будучи применена поочередно к натуральным числам $m = 0, 1, 2, 3, 4, 5, \ldots$, производит члены рассматриваемой последовательности.) Обратите внимание на следующие два факта относительно этой таблицы. Во-первых, *любая* вычислимая последовательность натуральных чисел должна появиться где-то (может быть, далеко не сразу) среди рядов таблицы. Это свойство выполнялось уже и для исходной таблицы, содержавшей значения □. Мы просто *добавили* несколько рядов, чтобы заменить «фиктивные» машины Тьюринга (т. е. такие, которые приводят к □ хотя бы в одном случае). Во-вторых, считая, что машина Тьюринга H существует, мы получили таблицу *вычислительным путем* (т. е. с помощью некоторого определенного алгоритма), а именно, посредством процедуры $T_n(m) \times H(n;m)$. Иными словами, существует некая машина Тьюринга Q, применение которой к паре чисел (n, m)

дает значение соответствующей ячейки таблицы. Для этой машины числа n и m на ленте можно кодировать таким же образом, как и для H, т. е. мы имеем

$$Q(n;m) = T_n(m) \times H(n;m).$$

Воспользуемся теперь разновидностью остроумного и мощного приема, так называемого *диагонального процесса* Георга Кантора. (Мы познакомимся с оригинальным вариантом этого метода в следующей главе.) Рассмотрим значения в ячейках, расположенных на главной диагонали таблицы — диагональные элементы (матрицы), — выделенные **жирным** шрифтом:

0	0	0	0	0	0	0	0	0
0	**0**	0	0	0	0	0	0	0
1	1	**1**	1	1	1	1	1	1
0	2	0	**2**	0	2	0	2	0
1	1	1	1	**1**	1	1	1	1
0	0	0	0	0	**0**	0	0	0
0	0	1	0	2	0	**3**	0	4
0	1	2	3	4	5	6	**7**	8
0	1	0	0	1	0	0	0	**1**
.				.				.
.				.				.
.				.				.

Эти элементы образуют некоторую последовательность $0, 0, 1, 2, 1, 0, 3, 7, 1, \ldots$, к каждому члену которой мы теперь прибавим единицу:

$$1,\ 1,\ 2,\ 3,\ 2,\ 1,\ 4,\ 8,\ 2,\ \ldots\ .$$

Это, безусловно, механическая процедура, и, поскольку наша таблица была получена путем вычислений, мы получим новую вычислимую последовательность $1 + Q(n;n)$, т. е.

$$1 + T_n(n) \times H(n;n)$$

(с учетом того, что для диагональных элементов $n = m$). Но наша таблица содержит в себе *все* вычислимые последовательности, поэтому она должна содержать также и новую последовательность. Однако это невозможно! Ведь наша новая последовательность отличается от первого ряда первым элементом, от второго — вторым, от третьего — третьим, и т. д. Налицо явное противоречие, которое и устанавливает справедливость доказываемого нами утверждения о том, что машина Тьюринга H на самом деле не существует! Иными словами, *не существует универсального алгоритма для решения вопроса об остановке произвольной машины Тьюринга.*

Можно построить доказательство и по-другому. Для этого заметим, что из предположения о существовании H следует и существование машины Тьюринга с номером k, реализующей алгоритм (диагональный процесс!) $1 + Q(n;n)$, т. е. можно записать

$$1 + T_n(n) \times H(n;n) = T_k(n).$$

Но если мы подставим в это выражение $n = k$, то получится

$$1 + T_k(k) \times H(k;k) = T_k(k).$$

Мы приходим к противоречию, потому что если $T_k(k)$ останавливается, то мы имеем невыполнимое равенство

$$1 + T_k(k) = T_k(k)$$

(поскольку $H(k;k) = 1$), тогда как в случае безостановочного действия $T_k(k)$ (т. е. когда $H(k;k) = 0$) мы получаем не менее абсурдное соотношение

$$1 + 0 = \square.$$

Вопрос о том, останавливается ли конкретная машина Тьюринга или нет, представляет собой совершенно четко определенную математическую задачу (а ранее мы уже видели, что, наоборот, различные важные математические задачи могут быть сведены к вопросу об остановке машины Тьюринга). Таким образом, доказав, что не существует алгоритма для решения вопроса об остановке машины, Тьюринг показал (также как и Черч, который использовал свой собственный и весьма отличающийся подход), что не может быть и общего алгоритма для решения математических задач. *Проблема разрешимости* Гильберта не имеет решения!

Это не означает, что в каждом *отдельном* случае мы не в состоянии выяснить справедливость (или, наоборот, несостоятельность) некоторого конкретного математического утверждения или определить, остановится ли данная машина Тьюринга. С помощью интуиции, искусных технических приемов или же опираясь просто на здравый смысл, мы, вероятно, могли бы получить ответ на такие вопросы в частных случаях. (Так, например, если перечень

инструкций некоторой машины Тьюринга не включает *ни одной* команды STOP или же, наоборот, состоит *только* из таких команд, то одного здравого смысла достаточно для решения вопроса о ее остановке!) Но не существует ни одного алгоритма, который позволял бы решать *любую* математическую задачу или давал ответ на вопрос об остановке *любой* машины Тьюринга при любых вводимых в нее числах.

Может показаться, что мы пришли к выводу о существовании по-крайней мере *нескольких* неразрешимых математических вопросов. Однако это совсем не так! Мы *не* показали, что существует какая-то необычайно громоздкая машина Тьюринга, для которой (в некотором абсолютном смысле) *невозможно* решить вопрос об остановке при ее работе с каким-то особенно громоздким числом — в действительности, все как раз наоборот, как мы сможем скоро убедиться. Мы вообще ничего не говорили о неразрешимости какой-то *отдельной* задачи, а только лишь об *алгоритмической* неразрешимости *классов* задач. В каждом конкретном случае ответ будет либо «да», либо «нет», поэтому алгоритм для решения частной задачи, конечно, существует, а именно алгоритм, который при применении к этой задаче просто дает ответ «да» или, может быть, «нет»! Трудность в данном случае состоит в том, что мы не знаем, *какой* именно из имеющихся алгоритмов применять в том или ином случае. Это вопрос об установлении математической истинности отдельного утверждения, но не об общем решении проблемы для целого класса утверждений. Очень важно сознавать, что сами по себе алгоритмы не доказывают математическую истину. Решение о правомерности использования каждого алгоритма должно всегда приходить извне.

Как превзойти алгоритм

К вопросу о том, как установить истинность математических утверждений, мы вернемся позднее, в связи с теоремой Геделя (см. главу 4). Пока же я бы хотел обратить ваше внимание на то, что доказательство Тьюринга носит гораздо более конструктивный характер и не столь негативно, как могло показаться из предыдущего изложения. Мы ведь *не* показали, что есть некая определенная машина Тьюринга, для которой абсолютно невозможно решить, останавливается она или нет. Более того, если внимательно проследить за доказательством, то выяснится, что для кажущихся «чрезвычайно сложными» машин сама процедура Тьюринга, использованная для их построения, неявным образом *дает ответ*! Посмотрим, как это происходит. Допустим, у нас есть алгоритм, который *иногда* позволяет определить, что машина Тьюринга не остановится. Вышеописанная процедура Тьюринга позволяет *явно* проследить за вычислениями машины Тьюринга в случае, когда этот конкретный алгоритм не дает ответа на вопрос об остановке вычислительного процесса. Однако тем самым эта процедура дает *нам* в этом случае возможность узнать ответ! Конкретная машина Тьюринга, за работой которой мы следим, и вправду никогда не остановится.

Чтобы подробно разобраться в этом вопросе, предположим, что у нас есть некий алгоритм, который иногда позволяет решить проблему остановки. Как и ранее, мы обозначим этот алгоритм (машину Тьюринга) через H, но теперь мы допускаем, что этот алгоритм не всегда может точно определить, что машина Тьюринга не остановится:

$$H(n;m) = \begin{cases} 0 \text{ или } \Box, & \text{если } T_n(m) = \Box, \\ 1, & \text{если } T_n(m) \\ & \text{останавливается.} \end{cases}$$

так что $H(n;m) = \Box$ возможно в случае, когда $T_n(m) = \Box$. Существует немало алгоритмов типа $H(n;m)$. (Например, $H(n;m)$ мог бы просто давать на выходе $\mathbf{1}$, как только машина $T_n(m)$ останавливается, хотя *такой* алгоритм едва ли представляет большой практический интерес!)

Мы можем повторить процедуру Тьюринга, следуя уже пройденным путем, с той только разницей, что теперь некоторые из «$\Box$» останутся не замененными на нули. Как и ранее, применив диагональный процесс, получим

$$1 + T_n(n) \times H(n;n)$$

в качестве n-го элемента диагонали. (Мы будем иметь $\Box$ каждый раз, когда $H(n;n) = \Box$.

Отметим, что $\square \times \square = \square$, $1 + \square = \square$.) Это безупречно алгоритмизованное вычисление, поэтому оно может быть произведено некоторой машиной Тьюринга, скажем k-й, и тогда мы получим

$$1 + T_n(n) \times H(n; n) = T_k(n).$$

Для k-го диагонального элемента (т. е. $n=k$) мы имеем

$$1 + T_k(k) \times H(k; k) = T_k(k).$$

Если вычисления $T_k(k)$ останавливаются, то мы приходим к противоречию (в этом случае $H(k; k)$ должно равняться единице, но тогда возникнет невыполнимое равенство: $1+T_k(k) = T_k(k)$). Значит, $T_k(k)$ не может остановиться, т. е.

$$T_k(k) = \square.$$

Но алгоритм не может этого «знать», потому что, если бы он давал $H(k; k) = 0$, мы снова пришли бы к противоречию (мы получили бы тогда неверное соотношение $1+0=\square$).

Таким образом, если мы можем отыскать k, то мы знаем, как построить вычислительную процедуру, для которой алгоритм не дает решения проблемы остановки, но *нам* ответ известен! А как нам найти k? Это непростая задача. Необходимо тщательно изучить конструкцию $H(n; m)$ и $T_n(m)$ и понять, как в точности действует $1 + T_n(n) \times H(n; n)$ в качестве машины Тьюринга. Затем надо определить номер этой машины, который и есть k. Конечно, это выполнить трудно, но вполне возможно[8]. Из-за этих трудностей вычисление $T_k(k)$ нас бы вовсе не интересовало, не будь она специально предназначена для доказательства неэффективности алгоритма H! Важно то, что мы получили строго определенную процедуру, которая для любого наперед заданного алгоритма H позволяет найти такое k, что для $T_k(k)$ этот алгоритм не может решить проблему остановки, т. е. мы тем самым превзошли его. Возможно, мысль о том, что мы «умнее» каких-то алгоритмов, принесет нам некоторое удовлетворение!

[8] Фактически, самую трудную часть мы уже выполнили, когда построили универсальную машину Тьюринга U, поскольку она позволяет нам записывать $T_n(n)$ как машину Тьюринга, действующую на n.

На самом деле, упомянутая процедура настолько хорошо определена, что мы могли бы даже найти *алгоритм* для нахождения k по заданному H. Поэтому, прежде чем мы «погрязнем» в самодовольстве, мы должны осознать, что *этот* алгоритм может улучшить H [9], поскольку он, по сути, «знает», что $T_k(k) = \square$, — или все-таки нет? В предыдущем изложении было удобно использовать антропоморфный термин «знать» по отношению к алгоритму. Однако не мы ли в конечном счете «знаем», тогда как алгоритм просто следует определенным нами правилам? А может быть мы сами просто следуем правилам, запрограммированным в конструкции нашего мозга и в окружающей нас среде? Эта проблема затрагивает не только алгоритмы, но и то, как мы выносим суждения об истинности и ложности. К этим важнейшим проблемам мы вернемся позднее. Вопрос о математической истине (и ее неалгоритмической природе) будет рассмотрен в главе 4. На данный момент мы, по крайней мере, получили некоторое представление о *значении* слов «алгоритм» и «вычислимость» и достигли понимания некоторых из относящихся к ним вопросов.

Лямбда-исчисление Черча

Понятие вычислимости — очень важная и красивая математическая идея. Примечателен также и ее малый возраст в сравнении с другими столь же фундаментальными математическим проблемами: она была впервые выдвинута только в 1930-х годах. Эта проблема имеет отношение ко *всем* областям математики (хотя, справедливости ради, отметим, что большинство математиков пока не часто обращаются к вопросам вычислимости). Сила этой идеи связана отчасти с существованием четко определенных и все же *неразрешимых* математических операций (как, например, проблема остановки машины Тьюринга и некоторые другие, которые мы рассмотрим в главе 4). Если бы не было таких невычислимых объектов, то теория алгоритмической разрешимости не представляла бы особого интереса для математики. В конце концов, математики любят головоломки.

Задача о разрешимости определенной математической операции может их заинтриговать, особенно потому, что общее решение *этой* головоломки само по себе алгоритмически не разрешимо.

Следует сделать еще одно замечание. Вычислимость — это по-настоящему «абсолютная» математическая идея. Это абстрактное понятие, которое никак не зависит от какой-либо конкретной реализации в терминах «машин Тьюринга» в том виде, как я их описал выше. Как я уже указывал, нет необходимости придавать какое-либо специальное значение «лентам», «внутренним состояниям» и т. п., характерным для гениального, но тем не менее частного подхода Тьюринга. Существуют также и другие способы выражения идеи вычислимости, причем исторически первым было «лямбда-исчисление», предложенное американским логиком Алонзо Черчем совместно со Стивеном Клини. Процедура, предложенная Черчем, значительно отличалась от метода Тьюринга и была гораздо более абстрактна. Фактически, форма, в которой Черч изложил свою теорию, делала связь между ними и чем бы то ни было «механическим» совсем не очевидной. Главная идея, лежащая в основе процедуры Черча, абстрактна по своей сути — это математическая операция, которую сам Черч назвал «абстрагированием».

Мне кажется, что стоит привести краткое описание схемы Черча не только потому, что она подчеркивает математическую природу идеи вычислимости, не зависящую от конкретного понятия вычислительной машины, но и потому, что она иллюстрирует мощь абстрактных идей в математике. Читатель, не достаточно свободный в математике и не увлеченный излагаемыми математическими идеями как таковыми, скорее всего предпочтет сейчас перейти к следующей главе — и не утратит при этом нить рассуждений. Тем не менее я полагаю, что таким читателям будет небесполезно следовать за мной еще какое-то время и оценить чудесную по своей стройности и продуманности схему Черча (см. Черч [1941]).

В рамках этой схемы рассматривается «универсальное множество» различных объектов, обозначаемых, скажем, символами

$$a, b, c, d, \ldots, z, a', b', \ldots, z',$$
$$a'', b'', \ldots, z'', a''', \ldots, a'''', \ldots,$$

каждый из которых представляет собой математическую операцию, или *функцию.* (Штрихованные буквы позволяют создавать неограниченные наборы символов для обозначения таких функций.) «Аргументы» этих функций, т. е. объекты, на которые эти функции действуют, в свою очередь являются объектами той же природы, т. е. функциями. Более того, результат действия одной функции на другую (ее «значение») также представляет собой функцию. (Поистине, в системе Черча наблюдается замечательная экономия понятий.) Поэтому, когда мы пишем[9]

$$a = bc,$$

мы подразумеваем, что функция b, действуя на функцию c, дает в результате другую функцию a. В рамках этой схемы нетрудно сформулировать понятие функции двух или более переменных. Если мы хотим представить f как функцию двух переменных, скажем p и q, то мы можем просто написать

$$(fp)q$$

(что есть результат действия функции fp на функцию q). Для функции трех переменных можно использовать выражение

$$((fp)q)r,$$

и так далее.

Теперь мы можем перейти к описанию важнейшей операции *абстрагирования.* Для нее мы будем использовать греческую букву λ (лямбда). Непосредственно за ней будет следовать символ одной из функций Черча, скажем x, который мы будем рассматривать как «фиктивную переменную». Каждое появление x в квадратных скобках, следующих сразу за этим выражением, обозначает теперь просто место, куда подставляется все,

[9] В более привычной форме эта запись имела бы вид $a = b(c)$, но эти дополнительные скобки в действительности не нужны, поэтому лучше просто привыкнуть к их отсутствию. Их последовательное использование привело бы к довольно громоздким формулам вида $(f(p))(q)$ и $((f(p))(q))(r)$ вместо $(fp)q$ и $((fp)q)r$, соответственно.

что идет за всем этим выражением. Таким образом, когда мы пишем

$$\lambda x.[fx],$$

мы подразумеваем функцию, которая при действии на, например, a имеет значение fa, т.е.

$$(\lambda x.[fx])a = fa.$$

Другими словами, $\lambda x.[fx]$ — это просто функция f, т.е.

$$\lambda x.[fx] = f.$$

Сказанное выше требует определенного осмысления. Это одна из тех математических тонкостей, которые на первый взгляд кажутся настолько педантичными и тривиальными, что их смысл часто совершенно ускользает от понимания. Рассмотрим пример из знакомой всем школьной математики. Примем за f тригонометрическую функцию — синус угла. Тогда абстрактная функция «sin» будет определяться выражением

$$\lambda x.[\sin x] = \sin .$$

(Не придавайте большого значения тому, что в качестве «функции» x может фигурировать величина угла. Мы скоро увидим, каким образом числа можно иногда рассматривать как функции, а величина угла — это просто число.) До сих пор все *на самом деле* тривиально. Однако представим себе, что обозначение «sin» не было изобретено, но нам известно о существовании представления $\sin x$ в форме степенного ряда:

$$x - \frac{1}{6}x^3 + \frac{1}{120}x^5 - \dots .$$

Тогда мы могли бы ввести определение

$$\sin = \lambda x.\left[x - \frac{1}{6}x^3 + \frac{1}{120}x^5 - \dots\right].$$

Можно было поступить еще проще и определить, например, операцию «одна шестая куба», для которой не существует стандартного «функционального» обозначения:

$$Q = \lambda x.\left[\frac{1}{6}x^3\right].$$

Тогда, например,

$$Q(a+1) = \frac{1}{6}(a+1)^3 = \frac{1}{6}a^3 + \frac{1}{2}a^2 + \frac{1}{2}a + \frac{1}{6}.$$

К обсуждаемым проблемам большее отношение имеют выражения, составленные просто из элементарных функциональных операций Черча, таких как

$$\lambda f.[f(fx)].$$

Это функция, которая, действуя на другую функцию, скажем g, дает дважды итерированную g, действующую на x

$$(\lambda f.[f(fx)])g = g(gx).$$

Мы могли бы сначала «абстрагироваться» от x и рассмотреть выражение

$$\lambda f.[\lambda x.[f(fx)]],$$

которое можно сократить до

$$\lambda fx.[f(fx)].$$

Это и есть операция, применение которой к g дает функцию «вторая итерация g». По сути, это та самая функция, которую Черч обозначил номером $\mathbb{2}$:

$$\mathbb{2} = \lambda fx.[f(fx)],$$

так что $(\mathbb{2}g)y = g(gy)$. Аналогичным образом он определил:

$$\mathbb{3} = \lambda fx.[f(f(fx))],$$
$$\mathbb{4} = \lambda fx.[f(f(f(fx)))], \quad \text{и т.д.,}$$

а также

$$\mathbb{1} = \lambda fx.[fx] \quad \text{и} \quad \mathbb{0} = \lambda fx.[x].$$

Видно, что $\mathbb{2}$ Черча больше похоже на «дважды», $\mathbb{3}$ — на «трижды» и т.д. Значит, действие $\mathbb{3}$ на функцию f, т.е. $\mathbb{3}f$ равносильно операции «применить f три раза», поэтому $\mathbb{3}f$ при действии на y превращается в $(\mathbb{3}f)y = f(f(f(y)))$.

Посмотрим, как в схеме Черча можно представить очень простую математическую операцию — прибавление $\mathbb{1}$ к некоторому числу. Определим операцию

$$S = \lambda abc.[b((ab)c)].$$

Чтобы убедиться, что S действительно прибавляет $\mathbb{1}$ к числу в обозначениях Черча, проверим ее действие на $\mathbb{3}$:

$$S\mathbb{3} = \lambda abc.[b((ab)c)]\mathbb{3} = \lambda bc.[b((\mathbb{3}b)c)] =$$
$$= \lambda bc.[b(b(b(bc)))] = \mathbb{4},$$

поскольку $(\mathbb{3}b)c = b(b(bc))$. Очевидно, эта операция с таким же успехом может быть применена к любому другому натуральному числу Черча. (В действительности, операция $\lambda abc.[(ab)(bc)]$ приводит к тому же результату, что и S.)

А как насчет удвоения числа? Удвоение числа может быть получено с помощью операции

$$D = \lambda abc.[(ab)((ab)c)],$$

что легко видеть на примере ее действия на $\mathbb{3}$:

$$D = \lambda abc.[(ab)((ab)c)]\mathbb{3} = \lambda bc.[(\mathbb{3}b)((\mathbb{3}b)c)] =$$
$$= \lambda bc.[(\mathbb{3}b)(b(b(bc)))] =$$
$$= \lambda bc.[b(b(b(b(b(bc)))))] = \mathbb{6}.$$

Фактически, основные арифметические операции — сложение, умножение и возведение в степень — могут быть определены, соответственно, следующим образом:

$$A = \lambda fgxy.[((fx)(gx))y],$$
$$M = \lambda fgx.[f(gx)],$$
$$P = \lambda fg.[fg].$$

Читатель может самостоятельно убедиться (или же принять на веру), что

$$(A\mathbb{m})\mathbb{n} = \mathbb{m} + \mathbb{n},$$
$$(M\mathbb{m})\mathbb{n} = \mathbb{m} \times \mathbb{n},$$
$$(P\mathbb{m})\mathbb{n} = \mathbb{n}^{\mathbb{m}},$$

где $\mathbb{m}$ и $\mathbb{n}$ — функции Черча для двух натуральных чисел, $\mathbb{m} + \mathbb{n}$ — функция, выражающая их сумму, и т. д. Последняя из этих функций поражает больше всего. Посмотрим, например, что она дает в случае $\mathbb{m} = \mathbb{2}$, $\mathbb{n} = \mathbb{3}$:

$$(P\mathbb{2})\mathbb{3} = ((\lambda fg.[fg])\mathbb{2})\mathbb{3} = (\lambda g.[\mathbb{2}g])\mathbb{3} =$$
$$= (\lambda g.[\lambda fx.[f(fx)]g])\mathbb{3} = \lambda gx.[g(gx)]\mathbb{3} =$$
$$= \lambda x.[\mathbb{3}(\mathbb{3}x)] = \lambda x.[\lambda fy.[f(f(fy))](\mathbb{3}x)] =$$
$$= \lambda xy.[(\mathbb{3}x)((\mathbb{3}x)((\mathbb{3}x)y))] =$$
$$= \lambda xy.[(\mathbb{3}x)((\mathbb{3}x)(x(x(xy))))] =$$
$$= \lambda xy.[(\mathbb{3}x)(x(x(x(x(x(xy))))))] =$$
$$= \lambda xy.[x(x(x(x(x(x(x(x(xy)))))))))] = \mathbb{9} = \mathbb{3}^{\mathbb{2}}.$$

Операции вычитания и деления определяются не так легко (на самом деле нам потребуется соглашение о том, что делать с $(\mathbb{m} - \mathbb{n})$, когда $\mathbb{m}$ меньше $\mathbb{n}$, и с $(\mathbb{m}/\mathbb{n})$, когда $\mathbb{m}$ не делится на $\mathbb{n}$). Решающий шаг в развитии этого метода был сделан в начале 1930-х годов, когда Клини удалось найти выражение для операции вычитания в рамках схемы Черча! Затем были описаны и другие операции. Наконец, в 1937 году Черч и Тьюринг независимо друг от друга показали, что всякая вычислимая (или алгоритмическая) операция — теперь уже в смысле машин Тьюринга — может быть получена в терминах одного из выражений Черча (и наоборот).

Это воистину замечательный факт, который подчеркивает глубоко объективный и математичный характер понятия вычислимости. На первый взгляд, понятие вычислимости по Черчу не связано с вычислительными машинами. И тем не менее, оно имеет непосредственное отношение к практическим аспектам вычислений. В частности, мощный и гибкий язык программирования LISP включает в себя как существенный элемент основные структуры исчисления Черча.

Как я отмечал ранее, существуют и другие способы определения понятия вычислимости. Несколько позже, но независимо от Тьюринга, Пост предложил во многом сходную концепцию вычислительной машины. Тогда же благодаря работам Дж. Хербранда и Геделя появилось и более практичное определение вычислимости (рекурсивности). Х. Б. Карри в 1929 году, и ранее, в 1924, М. Шенфинкель, предложили иной подход, который был отчасти использован Черчем при создании своего исчисления (см. Ганди [1988]). Современные подходы к проблеме вычислимости (такие как машина с неограниченным регистром, описанная Катлендом [1980]) в деталях значительно отличаются от разработанного Тьюрингом и более пригодны для практического использования. Однако *понятие* вычислимости во всех этих подходах остается неизменным.

Как и многие другие математические идеи, особенно наиболее фундаментальные и красивые, идея вычислимости кажется

овеществленной и объективно существующей в *платоновском* смысле. Именно к этому мистическому вопросу о платоновской реальности математических понятий в целом мы и обратимся в следующих двух главах.

Примечания

1. Я использую обычную современную терминологию, в которой множество «натуральных чисел» включает и нуль.

2. Существует немало других известных в математике способов записи пар, троек и большего количества чисел в виде одного числа, но они менее удобны для наших целей. Например, формула $(1/2)((a+b)^2+3a+b)$ однозначно представляет пару (a, b) как одно натуральное число. Проверьте сами!

3. В изложенном выше я не вводил никакой метки для начала последовательности чисел (или инструкций и т. п.). Это совершенно не требуется для входных данных, поскольку все начинается в тот момент, когда считана первая единица. Однако для конечного результата может понадобиться что-то дополнительное, поскольку *априори* никто не может сказать, как долго придется двигаться по ленте, чтобы добраться до первой (т. е. самой левой!) единицы. Хотя при движении налево может встретиться длинная строка нулей, нет никаких гарантий, что еще *дальше* не встретится единица. В этом случае применимы различные подходы. Можно было бы всегда использовать специальную отметку (допустим, 6, записанную при помощи процедуры «сокращения»), чтобы указывать начало и завершение окончательного ответа. Но для простоты я в своем изложении буду придерживаться другой точки зрения, согласно которой мы всегда «знаем», сколько в действительности ленты обработало наше устройство (например, можно представить, что оно оставляет своего рода «след»), так что не обязательно просматривать ленту до бесконечности, чтобы убедиться в том, что весь ответ считан.

4. Один из способов записи информации с двух лент на одну — вставить записи одной из них между записями другой. При этом нечетные отметки на новой ленте могут соответствовать отметкам первой ленты, тогда как четные — отметкам второй. Аналогичная схема работает и для четырех, и для большего числа лент. «Неэффективность» этой процедуры обусловлена тем, что считывающему устройству пришлось бы «прыгать» взад-вперед по ленте, оставляя на ней маркеры как на четных местах, так и на нечетных, с тем чтобы фиксировать свое положение в каждый момент.

5. Эта процедура имеет отношение только к методу, который позволяет интерпретировать запись на ленте как натуральное число. Она не изменяет номера наших конкретных машин Тьюринга, таких как EUC и XN + 1.

6. Если T_n определена *некорректно*, то U будет действовать так, как если бы число, отвечающее n, обрывалось сразу по достижении последовательности из четырех или более единиц в двоичной записи n. Остаток выражения будет считан уже как число m, после чего устройство начнет совершать некие бессмысленные вычисления! От этого свойства можно при желании избавиться, если представлять n в *расширенной* двоичной форме. Я решил не делать этого, чтобы еще больше не усложнять описание несчастной универсальной машины Тьюринга!

7. Я благодарен Давиду Дойчу за то, что он нашел десятичную форму двоичного представления u, которое я привожу ниже. Я признателен ему также за проверку того факта, что это двоичное значение u *действительно* задает универсальную машину Тьюринга! Двоичная запись u выглядит следующим образом:

   10000000010111010011010001001010101110100
   0110100010100000110101001101000101010010
   1101000011010001010010101101001001110100
   1010010010111010100011101010100100101011
   1010101001101000101000101011010000011010
   0100000101011010001001110100101000010101
   1101001000111010010101000010111010010100
   1101000010000111010100001110101000010010
   0111010001010101101010010101101000001101
   0101001011010010010001101000000001101000
   0001110101001010101011101000010011101001
   0101010101010111010000101010111010000101
   0001011101000101001101001000010100110100
   1010010011010010001011010100010111010010
   0101011101001010001110101001010010011101
   0101010000110100101010101110101001000101
   10101000001011010100010011010101010100010
   1101001010100100101101010010010111010101
   0010101110101001010011010101000011101000

1001001010111010101001010111010101000001
1101010010000011010101010010111010100101
0110100010010001110100000001110100101001
0101010111010010100100101011101000001010
1110100001000111010000010101001110100001
01001110100000010001011101000100001110100
0010010100111010001000010110100010100101
1101000101001011010010000010110100010101
00100110100010101010111010010000001110100
1001010101011101010101001101001000101011
01001001001011010000000101101000000100011
010000001001011010000000000110100101000101
1101001010100011010010100101011010000010
01110100101010010110100100111010100000001
0101110101000000110101010001010101101001
0101011010100001010111010100100101011101
01000100101101010010000101110100000001110
10100100001011010100101001101010100010111
0101001010010111010101000001011101010100
0001011101000000111010101000010101110100
1010101101010100001011101010001010101110
10101001001011101010101000011101010000000
0111010010010000110100100100010110101010
10100111010000000001011010010000110101010
1010010111010010000110100100010101011101
00001000111010001000011101000011010000000
01011010000010010111010101001010101110100
01000100101110100000100111010101001110100
0001010101101000010000111010010000100011
1010101010101001110100001001001110100010
0100001110100001010010110100001010000111
0101010101010111010001001001101000100100
1101010010100101110100010001010111010000
0001110100010010010111010011010010010000
1011010101010011010001010001011101000011
0101000010001011010100110101010010100101
1010101001101001001010111010011010010000
0101101000101010100011101001000010101101
0000001001101001000100101110100100001101
0100000100101110100100101001101001001010
1011010011010010010100101101001101001010
00001011010010000001110101001001101010101
0000101110100101000010111010010101010111
0101000100101101001001110100101010001011
1010001001110101000010110100100111010010
1010101011101001000111010010101010010111
01001000111010100000010101011100110101000
00101101001001110101000000010111010010110
1010000010101101001010010111010100001001
0111010000110101000100001011010100110101
0001000101101010101001011101010001010010
1101000101010101110100100001010110101000
1011101010010010101011101010100100101110
1010001110101000111010100100100101110101
0001110101001010001011101010001011101010
0001001011101010001110100010100010111010
0101001011101010010101001011101001010101
0101011010100001010101011010000100111010
0001010101010111010101000101011101010100
0101011101000000111010101000100101110100
0000111010101001000101110101000000110101
0000101101000000111010010000001011101010
0011101010010001010111010100110101010100
0101011010000011010101010010101011010000
0010011010101010010011101010011010101010
0100101101010011010010010011101000001101
0101010100101011010100010011010001010010
1010111010000011010101010101001011010001
0001110100010101010101011010001000111010
0001010111010001001000011101001101000000
0100111010000001001011101000100010100111
0100000010010111010010101010100101101000
0101010101110100010010100101110100000100
0101110101010010110100010001001110100000
10010101110100000001010101101000010001110
01111010000100000011101000010010011101000
00101001011101000000101001011010000100010
1011101000010001001101000100001110101111
01000010010010111010000100100101110100000
0001010111010000101010001101000100101110
10000100000011101000010011101000100000101
1101010100101101000100000101110100001010
10101110100000001010101110100010000101011
1010001000010101110100100000111010100100
1001101000000101011101000100010010111010
1010000111010100101011010010101010000110
1000001010011010000000111010000010010011
1010010110100100010100101101010100110100
0101001001011010101001101000101010001011
0011010100100101110101010011010001010101
0101100110101000101010110011010010001010
1010111010001000111010010010101010101101
0010100101000110100100000010111010000011
0101010010101010110100101010110100100010
0010111010001010101101010000010101101000
1000001101001000101011010000100111010100
1010101010111010010110100100100010101100
1101001001001010101110100110100100100101
0110100101101001001001001011010010110100
1001010001011001101001001010010101110100
0101011101001001011100110100100101010010
1110011010010100010101011101000100011101
0000101001011010010100010111010010100010
1011010001001110100101000100101110100010
0111010010100100010111001101001000100011
1010001001110100101001010101110011010010
1000001110011010101010101011010000000011101
0010101001010101110100100011101001010100
1010111001101000010100100110011010100000
1101000000011101001010101001010111001101
01000100001101000000001110100010010101010
11101000100011101010101010101010110100000
1001110100100010010101110100101010001001
10101000000001011010010011101010000101011
101001000011010100000000010110100100011101
0100100101110100001101010000101010110101

00010111010100001010010111010100010111010100010101010111001101010001010110100001101010001001010

Пытливый читатель, вооруженный эффективным домашним компьютером, быть может захочет проверить, используя данные в тексте предписания и применяя эту последовательность к номерам различных простых машин Тьюринга, что она и в самом деле соответствует универсальной машине Тьюринга!

Некоторое уменьшение величины *u* может быть достигнуто за счет другого определения машины Тьюринга. Например, мы могли бы отказаться от использования команды STOP и вместо этого применять правило остановки после того, как машина повторно возвращается во внутреннее состояние 0 из какого-либо другого внутреннего состояния. Это не дало бы нам значительного выигрыша (а может, и вовсе никакого). Большую пользу принесло бы использование лент с иными, нежели только **0** и **1**, отметками. В литературе встречаются описания очень компактных на вид машин Тьюринга, но эта компактность обманчива, поскольку она обусловлена чрезмерно сложным кодированием описаний машин Тьюринга вообще.

8. Желающие ознакомиться с вопросами, имеющими отношение к этому знаменитому утверждению и изложенными без излишних технических подробностей, могут обратиться к работе Дэвлина [1988].

9. Мы могли бы, конечно, «обыграть» и этот модифицированный алгоритм, просто за счет повторного применения предыдущей процедуры. Тогда мы сможем использовать эти вновь полученные знания для дальнейшего улучшения алгоритма, который мы, в свою очередь, снова превзойдем; и так далее. Тип рассуждений, в который выливается этот повторяющийся процесс, будет рассмотрен нами в связи с теоремой Геделя в главе 4 (с. 119).

Глава 3

Математика и действительность

Страна Тор'Блед-Нам

Представим себе, что мы совершаем большое путешествие в некий далекий мир. Назовем его Тор'Блед-Нам. Наша телеметрическая система зарегистрировала сигнал, вывела его на монитор и, отфокусировав изображение, мы увидели следующую картину (рис. 3.1):

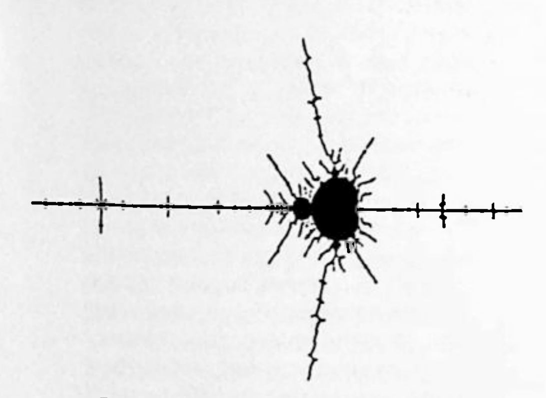


Рис. 3.1. Первый взгляд на новый мир

Что бы это могло быть? Странного вида насекомое? А может быть, темное озеро с многочисленными втекающими в него ручьями? Или огромный причудливой формы внеземной город, с исходящими в разных направлениях дорогами, которые ведут в расположенные поблизости городки и деревушки? Возможно, это остров — и если это так, то давайте поищем поблизости континент, с которым он связан. Для этого «отойдем назад», т. е. уменьшим увеличение наших приборов раз в 15. И вот — посмотрите-ка — этот новый мир предстал перед нашим взором во всей своей полноте (рис. 3.2):

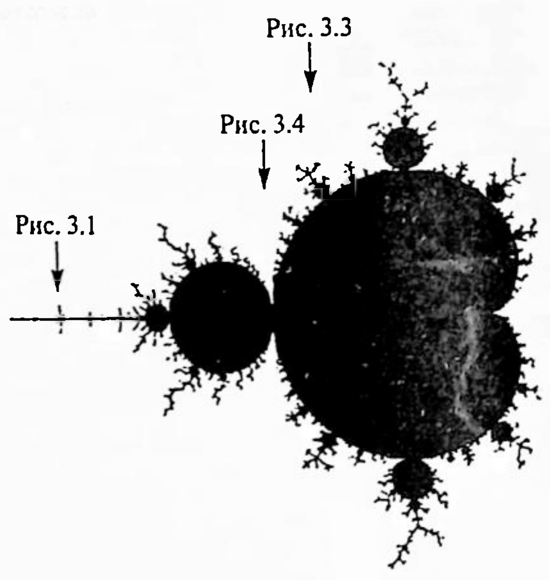


Рис. 3.2. Общий вид Тор'Блед-Нам. Стрелками отмечены области, увеличенные изображения которых даны на рис. 3.1, 3.3 и 3.4

На рис. 3.2 наш «островок» выглядит как маленькая точка под стрелкой «рис. 3.1». Все волокна (ручьи, дороги, мосты?), исходящие из первоначального островка, обрываются, за исключением одного — того, что выходит из внутренней части расположенной справа расщелины, и который, в свою очередь, соединен с объектом гораздо большего размера (он изображен на рис. 3.2). Последний, как нетрудно заметить, подобен

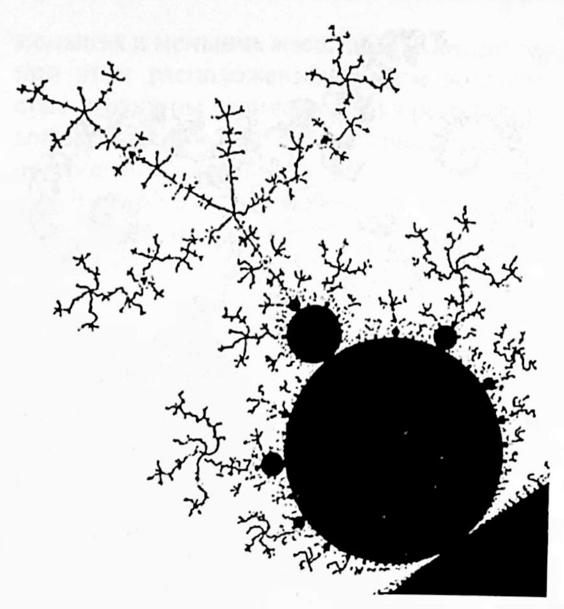


Рис. 3.3. Бородавка с «пятеричностью» своих волоконцев

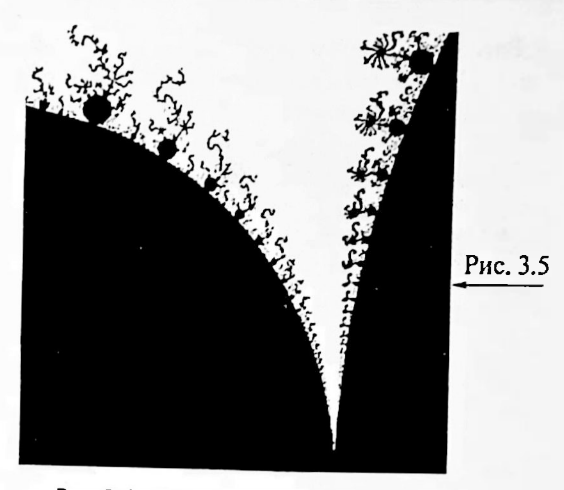


Рис. 3.4. Главная впадина. «Долина морских коньков» едва различима справа внизу

первоначальному островку, хотя их формы несколько отличаются. При более подробном рассмотрении «береговой линии» выявляются бесчисленные округлые выступы, края которых, в свою очередь, густо усеяны выступами такой же формы. Каждый маленький выступ соединен в каком-нибудь месте с более крупным, и все вместе они образуют бородавчатую структуру, где более крупные выступы покрыты наростами помельче, те — еще более мелкими и т. д. По мере того, как картина становится все более отчетливой, мы видим мириады мельчайших волокон, исходящих из рассматриваемой структуры. Сами волоконца ветвятся в разных местах, беспорядочно извиваясь. В некоторых частях волокон просматриваются узлы более сложной структуры, неразрешимые при данном увеличении приборов. Ясно, что наш объект — это никакой не остров или континент, и даже не пейзаж. Не исключено, что перед нашим взором чудовищный жук, а то, что мы увидели вначале, — это его детеныш, все еще соединенный с родителем своеобразной волокнистой пуповиной.

Давайте исследуем один из наростов у нашего насекомого, для чего увеличим разрешение примерно в десять раз (см. рис. 3.3 — соответствующая область на рис. 3.2. отмечена как «рис. 3.3»). Своим видом нарост сильно напоминает все существо целиком, за исключением места соединения. Обратите внимание, что на рис. 3.3 имеется множество точек, в которых сходятся пять волокон. По-видимому, этому конкретному наросту свойственна некая «пятеричность» (точно также как для самой верхней «бородавки» на рис. 3.2 характерна определенная «троичность»). На самом деле, если исследовать (на рис. 3.2) расположенный чуть ниже и левее следующий разумного размера нарост, то мы обнаружим у него «семеричность», а у следующего — характерную «девятеричность» и т. д. При углублении во впадину между двумя самыми крупными областями на рис. 3.2, справа будут встречаться наросты с постоянно нарастающим нечетным числом лучей. Давайте всмотримся внимательно вниз вглубь заостренной впадины, повысив увеличение еще в десять раз по сравнению с рис. 3.2 (рис. 3.4). Мы обнаружим множество других мельчайших наростиков на фоне общего беспорядочного завихрения. Справа видны едва различимые спиралевидные структуры, напоминающие «хвосты морских коньков», расположенные в области, которую мы так и назовем — «долина морских коньков». Здесь нам встретятся — если смотреть на это место при достаточно большом увеличении — разнообразные «морские анемоны» или области с богатой флорой. В конце концов, перед нами действительно может быть какой-то экзотический берег — возможно, коралловый риф, изобилующий

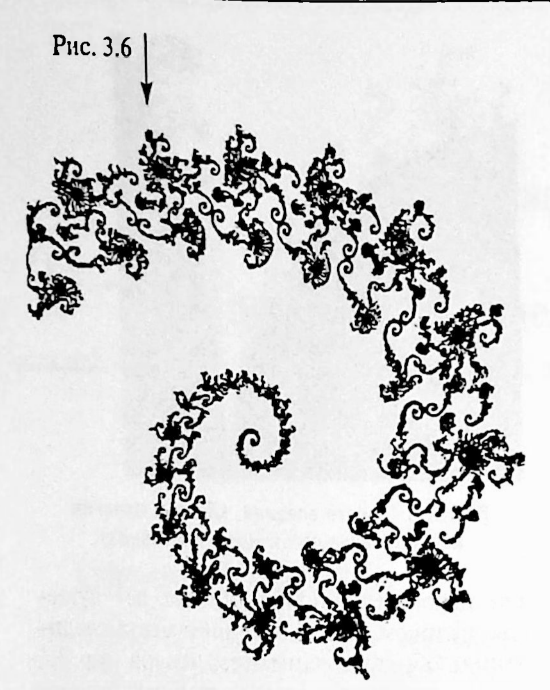


Рис. 3.5. Хвост «морского конька» крупным планом

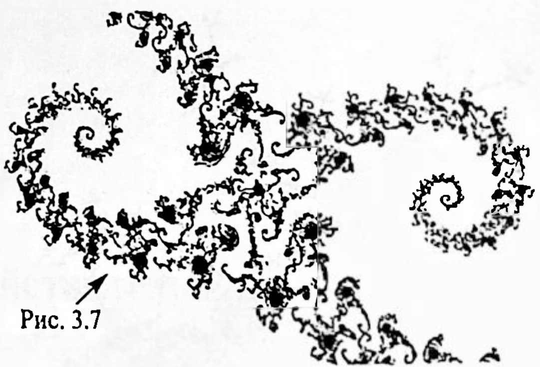


Рис. 3.6. Дальнейшее увеличение места соединения спиралей. В центре едва различим маленький детеныш

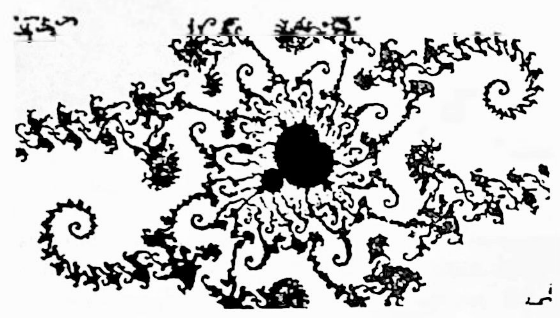


Рис. 3.7. При увеличении детеныш обнаруживает сходство с целым миром

всевозможными формами жизни. Объект, принятый нами за цветок, при более сильном увеличении может оказаться состоящим из мириада мельчайших и при этом невероятно сложных структур, с многочисленными волокнами и вихреобразными спиралевидными хвостами. Давайте рассмотрим подробнее один из более крупных хвостов морских коньков, а именно — едва различимое образование, обозначенное на рис. 3.4 как «рис. 3.5» (и соединенное с 29-ричным наростом!). Повысив увеличение в 250 раз, мы увидим изображенную на рис. 3.5 спираль. При этом окажется, что это не обычный хвост: и он тоже состоит из сложнейших вихреобразных структур с многочисленными мельчайшими спиралями и областями в форме осьминогов и морских коньков!

Во многих местах видно, что исследуемые нами структуры расположены точно в том месте, где сходятся две спирали. Рассмотрим одно такое место (обозначенное как «рис. 3.6» на рис. 3.5) с дополнительным 30-кратным увеличением. Посмотрите-ка: в самой середине теперь виднеется странный объект, в котором, однако, есть что-то знакомое. Увеличим изображение еще в шесть раз (рис. 3.7) — появляется крохотный дочерний объект, практически идентичный всей структуре! При более внимательном рассмотрении обнаруживаются некоторые отличия присоединенных к этой субструктуре волокон от тех, что выходят из основной структуры, — новые волокна, закручиваясь, уходят на значительно бóльшие относительные расстояния. И при этом маленькое существо выглядит почти неотличимым от своего родителя, — у него даже есть аналогично расположенные собственные детеныши. Можно было бы исследовать и их, если вновь повысить увеличение приборов. «Внуки» тоже будут напоминать своего общего предка — и нетрудно увидеть, что так может продолжаться до бесконечности. Этот странный мир Тор'Блед-Нам можно исследовать как угодно долго, постоянно увеличивая разрешающую способность нашей системы наблюдения. И тогда перед нами предстанет бесконечное разнообразие: никакие две области не являются в точности одинаковыми, но всем им свойственны общие черты, которые очень быстро становятся узнаваемыми. Знакомые нам уже жукообразные существа появляются на все

меньших и меньших масштабах. Каждый раз при этом расположенные рядом волокнистые структуры отличаются от предыдущих, демонстрируя новые фантастические сцены невероятной сложности.

В какой же странной и удивительно замысловатой по своей структуре стране мы оказались? Не сомневаюсь, что многие читатели уже знакомы с ней, но не все. Это не что иное, как фрагмент абстрактной математики — множество, известное под названием *множества Мандельброта*[1]. При всей его несомненной сложности оно получается на редкость простым образом! Чтобы как следует объяснить правила построения этого множества, необходимо сначала рассказать о том, что такое *комплексные числа*. Именно этим я сейчас займусь. Комплексные числа нам понадобятся и в дальнейшем. Они являются неотъемлемой частью структуры квантовой механики и вследствие этого лежат в основе поведения самого мира, в котором мы живем. Кроме того, комплексные числа являют собой одно из великих чудес математики. Чтобы объяснить, что такое комплексные числа, мне сначала потребуется напомнить вам, что подразумевается под термином «действительные числа». Не лишним будет также отметить связь этого понятия с действительностью «реального мира»!

Действительные числа

Напомним, что *натуральные* числа являются целыми величинами:

$$0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10,\ 11,\ \ldots .$$

Это самый элементарный и фундаментальный вид чисел. Ими можно количественно измерить любую дискретную сущность: можно говорить о двадцати семи овцах в поле, двух вспышках молнии, двенадцати ночах, тысяче слов, четырех беседах, нуле новых идей, одной ошибке, шести отсутствующих, двукратной смене направления и т. д. Натуральные числа можно складывать или перемножать, получая при этом новые натуральные числа. Мы использовали эти числа при обсуждении алгоритмов в предыдущей главе.

Тем не менее некоторые важные математические операции могут все же вывести нас за пределы мира натуральных чисел. Простейшая из них — вычитание. Для систематического определения вычитания нам понадобятся *отрицательные* числа. Теперь мы можем выстроить всю систему *целых чисел*:

$$\ldots -6,\ -5,\ -4,\ -3,\ -2,\ -1,\ 0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ \ldots .$$

Некоторые вещи — такие, как электрический заряд, банковские балансы или даты[1)], измеряются количественно этими числами. Однако сфера применения целых чисел все же слишком ограничена, поскольку *деление* одного числа на другое может оказаться неразрешимой задачей в рамках целых чисел. Соответственно, нам понадобятся *дроби*, или, как их называют, *рациональные числа*:

$$0,\ 1,\ -1,\ \frac{1}{2},\ -1,\ 2,\ -2,\ \frac{3}{2},\ -\frac{3}{2},\ \frac{1}{3},\ \ldots$$

Этих чисел достаточно для операций конечной арифметики, но для очень многих задач нам потребуется пойти еще дальше, с тем чтобы охватить бесконечные операции или операции перехода к пределу. Например, хорошо известная — и играющая огромную роль в математике — величина π возникает как результат многих бесконечных выражений. В частности, мы имеем:

$$\pi = 2\left(\frac{2}{1}\times\frac{2}{3}\times\frac{4}{3}\times\frac{4}{5}\times\frac{6}{5}\times\frac{6}{7}\times\frac{8}{7}\times\frac{8}{9}\times\ldots\right),$$

а также

$$\pi = 4\left(1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\frac{1}{9}-\frac{1}{11}+\ldots\right).$$

Это знаменитые выражения. Первое из них было найдено английским математиком, филологом и криптографом Джоном Уоллисом в 1655 году, а второе — шотландским математиком и астрономом (а также изобретателем первого телескопа-рефлектора) Джеймсом Грегори в 1671 году. Как и π, определенные подобным образом числа *не* обязаны быть рациональными (то есть представляться в виде m/n, где m и n — целые числа, причем n не равно нулю). Систему чисел

1) На самом деле при счете дат имеет место некоторое отступление от этого правила, поскольку нулевой год пропускается.

необходимо *расширить*, обеспечив возможность включения в нее таких величин.

Расширенная таким образом система чисел называется системой *действительных чисел* — тех самых хорошо знакомых нам чисел, что представляются в виде бесконечных десятичных дробей, таких как:

$$-583{,}70264439121009538\ldots .$$

В этом представлении мы получаем следующее известное выражение для числа π:

$$\pi = 3{,}14159265358979323846\ldots .$$

Другими примерами чисел, представимых таким образом, являются квадратные корни (или кубические корни, или корни четвертой степени) из положительных рациональных чисел, такие как:

$$\sqrt{2} = 1{,}41421356237309504\ldots$$

или же квадратные корни (или кубические корни и т. д.) любого положительного числа, как, например, выражение для числа π, найденное великим швейцарским математиком Леонардом Эйлером:

$$\pi = \sqrt{6\left(1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{25} + \frac{1}{36} + \ldots\right)}.$$

Действительные числа нам в сущности хорошо знакомы — мы с ними сталкиваемся в повседневной жизни. Правда обычно нас интересуют всего лишь приближения к этим числам и мы предпочитаем ограничиваться разложениями, состоящими из небольшого числа десятичных знаков. Тем не менее, в математических утверждениях может потребоваться *точное* задание действительных чисел и, как следствие, необходимость в некотором бесконечном способе описания наподобие бесконечной десятичной дроби, или какого-нибудь иного бесконечного математического выражения вроде приведенных выше формул для числа π, предложенных Уоллисом, Грегори и Эйлером. (В дальнейшем я буду обычно использовать десятичные дроби, но лишь потому, что они нам наиболее привычны. У математиков есть множество разных и более удовлетворительных способов представления действительных чисел, но нас это здесь не интересует.)

Может создаться впечатление, что представить себе все бесконечное десятичное разложение целиком невозможно, но это не так. Вот простой пример, когда вся последовательность знаков оказывается явным образом обозримой:

$$\frac{1}{3} = 0{,}333333333333333\ldots .$$

Многоточие указывает на то, что последовательность троек продолжается бесконечно. Для получения полного представления об этом разложении достаточно знать, что оно действительно состоит из неограниченной последовательности одних лишь троек. У каждого рационального числа есть повторяющееся (или конечное) десятичное представление вроде:

$$\frac{93}{74} = 1{,}2567567567567567\ldots ,$$

где последовательность 567 повторяется неограниченное число раз. Это число тоже оказывается полностью обозримым. Также обозримым является выражение

$$0{,}220002222000002222220000000222222220\ldots ,$$

которое определяет *иррациональное* число (оно просто состоит из последовательностей нулей и двоек, длины которых каждый раз увеличиваются на единицу), и еще много похожих выражений. В каждом таком случае нам достаточно знать правило, по которому составлено разложение. Знание алгоритма порождения очередной цифры в разложении числа — при условии, что такой алгоритм существует — дает нам способ «увидеть» целиком все бесконечное десятичное разложение. Действительные числа с алгоритмически порождаемыми десятичными разложениями называются *вычислимыми* числами (см. также с. 76). (При этом не важно, десятичное это разложение или двоичное. Вычислимыми в этом смысле оказываются одни и те же числа, независимо от использованного основания разложения.) Только что рассмотренные числа π и $\sqrt{2}$ представляют собой примеры вычислимых чисел. В обоих случаях подробное описание соответствующего правила — задача довольно-таки кропотливая, но, в принципе, нетрудная.

Есть, однако, действительные числа, которые *не* являются вычислимыми в упомянутом выше смысле. Как мы убедились

в главе 2, существуют невычислимые и при этом совершенно четко определенные последовательности. В качестве примера можно рассмотреть десятичное разложение, в котором n-я цифра равна 0 или 1 в зависимости от того, останавливается или нет n-я машина Тьюринга, производящая действия над числом n. В общем случае мы потребуем лишь, чтобы для действительного числа существовало *какое-нибудь* бесконечное десятичное разложение. Мы не только не требуем существования алгоритма порождения n-й цифры, но нам даже не обязательно знать о существовании какого бы то ни было правила, в принципе определяющего n-ю цифру [2]. Заметим, что вычислимые числа неудобны в работе. Невозможно обойтись одними лишь вычислимыми операциями, даже оперируя вычислимыми числами. Например, в общем случае вычислимым образом невозможно даже решить, равны ли два вычислимых числа друг другу! По этой причине мы будем работать со *всеми* действительными числами, когда десятичная последовательность может быть любой, а не только, скажем, вычислимой.

В заключение отметим также тождественность действительных чисел, чьи десятичные разложения заканчиваются бесконечной последовательностью девяток, и чисел, чьи разложения заканчиваются бесконечной последовательностью нулей. Например:

$$-27{,}1860999999\ldots = -27{,}1861000000\ldots.$$

Сколько же всего действительных чисел?

Давайте остановимся на минутку, чтобы оценить всю колоссальность обобщения при переходе от рациональных чисел к действительным.

Вначале может показаться, что целых чисел больше, чем натуральных, поскольку каждое натуральное число является целым, в то время как некоторые целые числа (а именно отрицательные) натуральными не являются. Аналогично может создаться впечатление, что дробей больше, чем целых чисел. Однако это не так. Согласно мощной и очень красивой теории бесконечных чисел, разработанной в конце XIX века Георгом Кантором — исключительно самобытным немецким математиком русского происхождения, — общее число дробных чисел, общее количество всех целых чисел и число всех натуральных чисел равны *одному и тому же* бесконечному числу, обозначаемому $\aleph_0$ («алеф-нуль»). (Удивительно, что похожая идея была частично предвосхищена еще за 250 лет до этого в начале XVII века великим итальянским физиком и астрономом Галилео Галилеем. Мы вспомним о некоторых других достижениях Галилея в главе 5.) Равенство количества целых чисел количеству натуральных чисел видно из следующего взаимно-однозначного соответствия:

Целые числа		Натуральные числа
0	↔	0
−1	↔	1
1	↔	2
−2	↔	3
2	↔	4
−3	↔	5
3	↔	6
−4	↔	7
.	.	.
.	.	.
.	.	.
$-n$	↔	$2n-1$
n	↔	$2n$
.	.	.
.	.	.
.	.	.

Обратите внимание, что каждое целое число (в левом столбце) и каждое натуральное число (в правом столбце) встречаются один и только один раз в своем списке. В канторовской теории множеств именно существование такого рода взаимно-однозначного соответствия устанавливает факт *равенства* числа объектов в левом столбце числу объектов в правом столбце. Таким образом, число целых чисел действительно равно числу натуральных чисел. В данном случае это число бесконечно, но это не имеет значения. (Единственное необычное свойство бесконечных чисел состоит в том, что даже если мы исключим некоторые элементы одного из списков, мы

можем установить взаимно-однозначное соответствие между элементами двух списков.) Аналогичным, хотя и несколько более сложным образом, устанавливается взаимно-однозначное соответствие между дробными и целыми числами. (Для этого можно использовать какой-либо из способов представления *пар* натуральных чисел — числителей и знаменателей — через отдельные натуральные числа; см. главу 2, с. 70.) Множества, которые можно поставить во взаимно-однозначное соответствие с рядом натуральных чисел, называются *счетными*; таким образом, счетные бесконечные множества — это множества, состоящие из $\aleph_0$ элементов. И, как мы только что убедились, множество целых чисел, равно как и множество дробных чисел, является счетным.

Существуют ли множества, *не* являющиеся счетными? Несмотря на расширение натуральной системы чисел сначала целыми, а затем и рациональными числами, общее число рассматриваемых объектов не увеличилось. Как мы убедились, число объектов во всех случаях осталось счетным. У читателя теперь может создаться впечатление, что *все* бесконечные множества счетны. Это не так, поскольку ситуация меняется коренным образом при переходе к действительным числам. Одним из замечательных достижений Кантора явилось доказательство того, что действительных чисел больше, чем натуральных. При этом Кантор применил так называемый *диагональный процесс*, который упоминался в главе 2 и который Тьюринг использовал в своем доказательстве неразрешимости *проблемы остановки* для машин Тьюринга. Доказательство Кантора, как и более позднее доказательство Тьюринга, — это *доказательство от противного*. Предположим, что утверждение, справедливость которого мы хотим установить, на самом деле ложно, то есть множество действительных чисел счетно. Тогда множество действительных чисел в интервале от 0 до 1 должно быть заведомо счетным и должен существовать *какой-нибудь* список, устанавливающий взаимно-однозначное соответствие между рассматриваемым множеством действительных чисел и множеством натуральных чисел, наподобие вот этого:

Натуральные числа		Действительные числа
0	↔	0,10357627183...
1	↔	0,14329806115...
2	↔	0,02166095213...
3	↔	0,43005357779...
4	↔	0,92550489101...
5	↔	0,59210343297...
6	↔	0,63667910457...
7	↔	0,87050074193...
8	↔	0,04311737804...
9	↔	0,78635081150...
10	↔	0,40916738891...
.	.	.
.	.	.
.	.	.

Жирным шрифтом выделены диагональные десятичные знаки. В данном случае эти цифры равны:

1, 4, 1, 0, 0, 3, 1, 4, 8, 5, 1, ...

Метод диагонального процесса состоит в построении действительного числа (в интервале от 0 до 1), чье десятичное разложение (после десятичной запятой) отличается в каждом разряде от соответствующего числа приведенной выше последовательности. Для определенности положим, что цифра данного разряда равна 1, если цифра соответствующего разряда на диагонали отлична от 1, и равна 2, если цифра на диагонали равна 1. Таким образом, в рассматриваемом случае получается такое действительное число:

0,21211121112 ... .

Это действительное число не может быть в списке, поскольку оно отличается от первого числа в первом десятичном разряде (после десятичной запятой), от второго числа — во втором разряде, от третьего числа — в третьем разряде и т. д. Таким образом, мы приходим к противоречию, поскольку полагали, что рассматриваемый список содержит *все* действительные числа в интервале от 0 до 1. Из этого противоречия следует истинность утверждения, которое нам требовалось доказать, — а именно, что не существует взаимно-однозначного соответствия между множеством действительных чисел и множеством натуральных чисел и, соответственно, что число действительных чисел *больше* числа рациональных чисел и *не* является счетным.

Число действительных чисел равно бесконечному числу, обозначаемому C. (Здесь C является сокращенным обозначением слова континуум — другого названия системы действительных чисел.) Может возникнуть вопрос, почему мы не обозначаем это число, например, $\aleph_1$. Символ $\aleph_1$ на самом деле обозначает следующее за $\aleph_0$ бесконечное число, а вопрос о том, верно ли утверждение $C = \aleph_1$ — это так называемая *континуум-гипотеза*, — представляет собой знаменитую и пока что нерешенную проблему.

При этом следует отметить, что множество *вычислимых* чисел счетно. Пересчитать их можно просто перечислив по порядку машины Тьюринга, порождающие действительные числа (то есть машины, последовательно порождающие цифры каждого разряда действительных чисел). При этом можно исключить из списка любую машину Тьюринга, порождающую действительное число, которое уже встречалось ранее в списке. Поскольку множество машин Тьюринга счетно, то, следовательно, счетным также должно быть и множество вычислимых действительных чисел. Почему же нельзя применить диагональный процесс к этому списку с тем, чтобы породить новое *не* включенное в список вычислимое число? Ответ состоит в том, что в общем случае невозможно с помощью вычислений решить, следует ли ту или иную машину Тьюринга включать в список, поскольку для этого мы должны были бы иметь возможность решить проблему остановки. Некоторые машины Тьюринга, начав порождение цифр действительного числа, могут зависнуть и оказаться уже не в состоянии выдать очередную цифру (поскольку они «не остановятся»). Не существует вычислимого способа, который позволил бы решить, какие именно машины Тьюринга зависнут таким образом. Это, в сущности, и есть проблема остановки. Значит, хотя метод диагонального процесса и породит некоторое действительное число, последнее не будет вычислимым. На самом деле, это рассуждение может использоваться *для доказательства* существования невычислимых чисел. Именно в этом ключе выдержано описанное в предыдущей главе тьюринговское доказательство существования классов алгоритмически неразрешимых задач. Другие области применения диагонального процесса будут рассмотрены дальше.

«Действительность» действительных чисел

Если отвлечься от понятия вычислимости, то действительные числа называются «действительными», потому что они, как представляется, дают величины, необходимые для измерения расстояний, углов, времени, энергии, температуры и многих других геометрических и физических параметров. Однако связь абстрактно определенных «действительных» чисел с физическими величинами не так проста, как может показаться. Действительные числа следует рассматривать скорее как некоторую *математическую идеализацию*, чем как реальную меру физически объективных величин. Система действительных чисел обладает, например, таким свойством, что между любыми двумя действительными числами (вне зависимости от их близости) существует третье действительное число. При этом совершенно не ясно, можно ли обоснованно утверждать то же самое о физических расстояниях или промежутках времени. Если мы продолжим дробить физическое расстояние между двумя точками, то мы в конце концов достигнем масштабов столь малых, что само понятие расстояния в обычном его смысле станет бессмысленным. Предполагается, что это действительно имеет место на масштабах, характерных для квантовой теории гравитации, которые в 10^{20} раз[2] меньше размеров субатомных частиц. Но чтобы отобразить действительные числа нам потребуется дойти до сколь угодно более мелких масштабов, которые, например, в 10^{200}, 10^{2000} или даже в $10^{10^{200}}$ раз меньше размеров частиц. И совершенно не ясно, есть ли какой бы то ни было физический смысл у столь абсурдно малых масштабов. То же самое можно сказать и в отношении столь же малых интервалов времени.

[2] Напомним, что 10^{20} означает число

100 000 000 000 000 000 000,

то есть единицу с двадцатью нулями.

Система действительных чисел выбрана в физике в силу ее *математической* полезности, простоты и изящества, а также поскольку она согласуется на очень широком интервале масштабов с физическими понятиями пространства и времени. Она выбрана *не* потому, что мы будто бы знаем, что она согласуется с упомянутыми физическими величинами на *всех* масштабах. Такое согласие вполне может не иметь места на очень малых пространственных и временны́х масштабах. Обычные расстояния измеряются при помощи линейки, но линейка оказывается «зернистой» при переходе к масштабам образующих ее атомов. Само по себе это не мешает нам продолжать использовать действительные числа подходящим образом, но измерение меньших расстояний требует уже гораздо большей изобретательности. По крайней мере, мы должны быть готовы предположить, что на очень-очень малых масштабах могут встречаться принципиальные трудности с расстояниями. Как оказывается, природа оказалась к нам на удивление благосклонна, сделав те самые действительные числа, которые мы привыкли повседневно применять для описания предметов на макро-масштабах, пригодными для описания расстояний гораздо меньших атомных — по крайней мере, на масштабах, равных одной сотой «классического» диаметра элементарной частицы — такой, как электрон или протон, — и, по-видимому, вплоть до «масштабов квантовой теории гравитации», что на двадцать порядков меньше размеров таких частиц! Это пример исключительно сильной экстраполяции нашего опыта. Сфера применимости привычного понятия расстояния, измеряемого действительными числами, по-видимому, простирается до самых далеких квазаров и еще дальше. Общий диапазон измеримых расстояний составляет 10^{42}, а может быть, 10^{60} или даже больше. Кстати, сомнения в правомерности использования системы действительных чисел высказывались не так уж часто. Почему же мы так уверены в том, что эти числа дают точное описание физических явлений, хотя реально об их применимости мы знаем лишь в весьма ограниченном диапазоне масштабов? Должно быть, эта уверенность — возможно, неверная — основывается на (правда, не очень часто признаваемых) логическом изяществе, внутренней согласованности и математической мощи системы действительных чисел в сочетании с верой в глубинную математическую гармонию природы.

Комплексные числа

Оказывается, что действительные числа — это не единственная математически мощная и изящная система чисел. Система действительных чисел все же не лишена некоторых неудобств. Например, квадратные корни можно извлекать только из положительных чисел (или нуля), но никак не из отрицательных чисел. С математической точки зрения — и отвлекаясь пока что от вопроса о непосредственной связи с физическом миром — было бы очень удобно иметь возможность извлекать квадратные корни как из положительных, так и из отрицательных чисел. Давайте постулируем существование, или попросту «изобретем» квадратный корень из числа -1. Обозначим его буквой i. Тогда мы имеем:

$$i^2 = -1.$$

Величина i, конечно же, не может быть действительным числом, поскольку произведение действительного числа на самого себя всегда положительно (или равно нулю, если само число равно нулю). Поэтому числа, квадраты которых отрицательны, обычно называют *мнимыми*. Следует, однако, отметить, что эти «мнимые» числа не менее реальны, чем ставшие уже привычными «действительные» числа. Как я уже отмечал выше, связь таких «действительных» чисел с *физической* реальностью далеко не столь непосредственна и убедительна, как может показаться на первый взгляд, и основана на математической идеализации о допустимости бесконечного уточнения, которая не имеет ясного *априорного* обоснования в природе.

Имея квадратный корень из -1, можно без особого труда получить квадратные корни для всех действительных чисел. Если a является положительным действительным числом, то величина

$$i \times \sqrt{a}$$

есть квадратный корень из отрицательного действительного числа $-a$. (У этого числа есть еще другой квадратный корень, а именно $-i \times \sqrt{a}$.) Ну, а что же можно сказать о самом числе i? Есть ли у него квадратный корень? Разумеется есть, поскольку, как легко проверить, величина

$$\frac{1+i}{\sqrt{2}}$$

(равно как и та же величина, взятая с отрицательным знаком), будучи возведена в квадрат, равна i. А у этой величины, в свою очередь, есть квадратный корень? Ответ опять положительный: квадрат числа

$$\sqrt{\frac{1+1/\sqrt{2}}{2}}+i\sqrt{\frac{1-1/\sqrt{2}}{2}}$$

или того же числа, взятого с отрицательным знаком, действительно равен $(1+i)/\sqrt{2}$.

Обратите внимание, что при образовании такого рода величин мы позволили себе складывать действительные и мнимые числа, а также умножать наши числа на произвольные действительные числа (или делить их на произвольные ненулевые действительные числа, а это то же самое, что умножать их на обратные величины). Получаемые таким образом объекты называются *комплексными числами*. Комплексное число это число вида:

$$a+ib,$$

где a и b — это действительные числа, называемые, соответственно, *действительной* и *мнимой частью* комплексного числа. Правила сложения и умножения двух таких чисел вытекают из обычных правил (школьной) алгебры с одним дополнительным правилом $i^2=-1$:

$$(a+ib)+(c+id)=(a+c)+i(b+d),$$
$$(a+ib)\times(c+id)=(ac-bd)+i(ad+bc).$$

Удивительное дело: к созданию этой системы чисел нас подтолкнуло желание иметь возможность извлечения квадратных корней из любых чисел. Эта цель достигнута, хотя само по себе это еще не очевидно. Но новая система чисел позволяет делать гораздо больше: безнаказанно извлекать кубические корни, корни пятой степени, корни девяносто девятой степени, корни π-й степени, корни степени $1+i$ и т.д. (это смог доказать еще в XVIII веке великий математик Леонард Эйлер). В качестве другого примера волшебных свойств комплексных чисел рассмотрим довольно сложные на вид тригонометрические формулы, которые проходят в школе. Так, синус и косинус суммы двух углов

$$\sin(A+B)=\sin A\cos B+\cos A\sin B,$$
$$\cos(A+B)=\cos A\cos B-\sin A\sin B$$

представляют собой, соответственно, просто-напросто мнимую и действительную части гораздо более простого (и легче запоминаемого!) комплексного уравнения[3]:

$$e^{iA+iB}=e^{iA}e^{iB}.$$

Все, что нам нужно здесь знать, это «формула Эйлера» (по-видимому, полученная за много лет до Эйлера замечательным английским математиком XVI века Роджером Котсом):

$$e^{iA}=\cos A+i\sin A,$$

которую мы теперь подставим в приведенное выше уравнение. В результате имеем:

$$\cos(A+B)+i\sin(A+B)=$$
$$=(\cos A+i\sin A)(\cos B+i\sin B),$$

и, выполнив умножение в правой части, получим искомые тригонометрические соотношения.

Более того, любое алгебраическое уравнение

$$a_0+a_1z+a_2z^2+a_3z^3+\ldots+a_nz^n=0$$

(где $a_0, a_1, a_2, \ldots, a_n$ являются комплексными числами и $a_n\neq 0$) всегда имеет своим решением некоторое комплексное число z. Например, существует комплексное число, удовлетворяющее соотношению:

$$z^{102}+999z^{33}-\pi z^2=-417+i,$$

хотя это совершенно не очевидно!

[3] Величина $e=2{,}7182818285\ldots$ (основание натуральных логарифмов, иррациональное число, по своему значению для математики сравнимое с числом π) определяется как

$$e=1+\frac{1}{1}+\frac{1}{1\times 2}+\frac{1}{1\times 2\times 3}+\ldots,$$

а e^z означает степень z от e, которая равна

$$e^z=1+\frac{z}{1}+\frac{z^2}{1\times 2}+\frac{z^3}{1\times 2\times 3}+\ldots.$$

Это общее свойство иногда называют «основной теоремой алгебры». Многие математики XVIII века старались доказать этот результат. Получить удовлетворительное доказательство в общем случае оказалось не под силу даже Эйлеру. И только в 1831 году великий математик и естествоиспытатель Карл Фридрих Гаусс предложил потрясающий по своей оригинальности ход рассуждений и представил первое общее доказательство. Ключевым компонентом этого доказательства было применение топологических[4)] рассуждений к *геометрическому* представлению комплексных чисел.

На самом деле Гаусс не был первым, кто использовал геометрическое представление комплексных чисел. Уоллис сделал то же самое примерно за двести лет до Гаусса, хотя далеко не столь результативно. Геометрическое представление комплексных чисел обычно связывают с именем Жана Робера Аргана — швейцарского бухгалтера, описавшего это представление в 1806 году, хотя полное описание этого представление было на самом деле дано девятью годами раньше норвежским геодезистом Каспаром Весселем. Согласно этой традиционной (хотя и не совсем правильной с исторической точки зрения) терминологии, я буду называть стандартное геометрическое представление комплексных чисел плоскостью Аргана.

Плоскость Аргана представляет собой обычную евклидову плоскость со стандартными декартовыми координатами x и y, где x обозначает расстояние по горизонтали (положительное вправо и отрицательное влево), а y — расстояние по вертикали (положительное вверху и отрицательное внизу). В этом случае комплексное число

$$z = x + iy$$

представляется точкой на плоскости Аргана с координатами

$$(x, y)$$

(рис. 3.8). Обратите внимание, что число 0 (рассматриваемое как комплексное число)

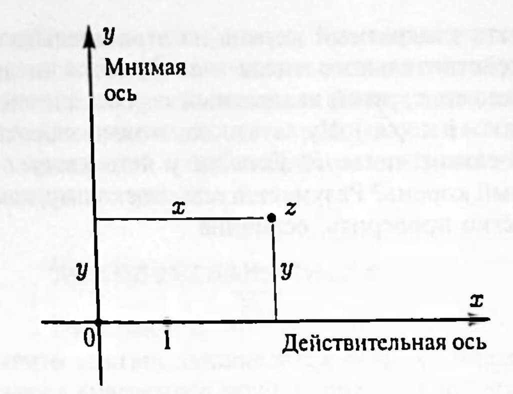


Рис. 3.8. Изображение комплексного числа $z = x + iy$ на плоскости Аргана

соответствует началу координат, а число 1 — одной из точек на оси x.

Плоскость Аргана есть просто способ геометрически наглядной организации семейства комплексных чисел. Такое представление не является для нас чем-то совершенно новым. Мы уже знакомы с геометрическим представлением *действительных* чисел — в виде прямой линии, простирающейся на неограниченное расстояние в обоих направлениях. Одна из точек обозначена как 0, а еще одна — как 1. Точка 2 смещена относительно точки 1 равно настолько, насколько точка 1 смещена относительно точки 0; точка 1/2 расположена в точности посередине между точками 0 и 1; точка −1 расположена так, что точка 0 находится в точности посередине между точками −1 и 1, и т. д., и т. п. Отображенное таким образом множество действительных чисел называется *действительной прямой*. В случае комплексных чисел у нас есть уже целых два действительных числа — a и b — которые могут рассматриваться как координаты комплексного числа $a + ib$. Эти два числа дают нам две координаты точки на плоскости, в данном случае — на плоскости Аргана. Для примера я указал на рис. 3.9 приблизительные положения комплексных чисел

$$u = 1 + i1{,}3; \quad v = -2 + i; \quad w = -1{,}5 - i0{,}4.$$

Теперь основные алгебраические операции сложения и умножения комплексных чисел приобретают ясную геометрическую интерпретацию. Рассмотрим сначала сложение. Предположим, что u и v это два комплексных числа, представленные на плоскости Аргана в соответствии с описанной

[4)] Слово «топологический» означает, что речь идет о разделе геометрии, — иногда называемом «геометрией резиновой поверхности», — в котором расстояния не имеют никакого значения, а важны только свойства непрерывности объектов.



Рис. 3.9. Расположение чисел $u = 1 + i1{,}3$, $v = -2 + i$, $w = -1{,}5 - i0{,}4$ на плоскости Аргана

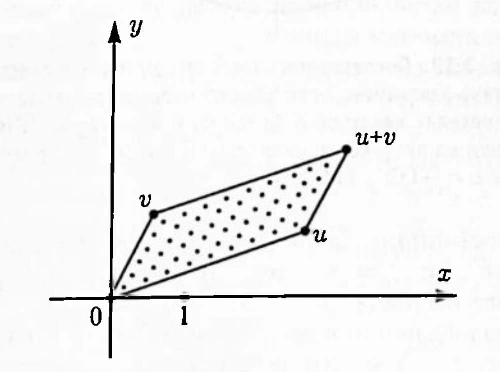


Рис. 3.10. Сумма $u + v$ двух комплексных чисел определяется по правилу параллелограмма

выше схемой. Тогда сумма этих двух чисел $u + v$ представляется «векторной суммой» двух точек, то есть точка $u + v$ находится на месте недостающей вершины параллелограмма, образованного точками u, v и началом координат 0. Нетрудно убедиться, что эта конструкция (рис. 3.10) действительно дает сумму двух чисел, но соответствующее доказательство я здесь опускаю.

Произведение uv двух комплексных чисел тоже имеет простую, хотя и, быть может, несколько менее очевидную геометрическую интерпретацию (рис. 3.11). (Я опять опускаю доказательство.) Угол при начале координат между 1 и uv равен сумме углов между 1 и v и между 1 и u (все углы измеряются против часовой стрелки), а расстояние точки uv от начала координат равно произведению расстояний от начала координат до u и v. Это эквивалентно утверждению, что треугольник, образованный точками 0, v и uv подобен (и ориентирован подобно)

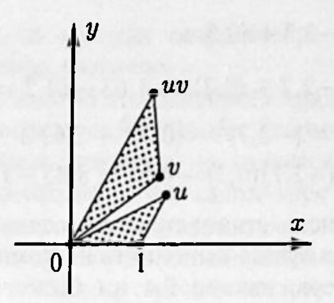


Рис. 3.11. Произведение uv двух комплексных чисел u и v — это такое число, что треугольник, образованный точками 0, v и uv, подобен треугольнику, образованному точками 0, 1 и u. То же самое можно сформулировать иначе: расстояние точки uv от 0 равно произведению расстояний от 0 до точек u и v, а угол между uv и действительной (горизонтальной) осью равен сумме углов между этой осью и отрезками к точкам u и v

треугольнику, образованному точками 0, 1 и u. (Энергичные читатели, не знакомые с такого рода построениями, могут сами убедиться в том, что эти построения непосредственно следуют из только что приведенных алгебраических правил сложения и умножения комплексных чисел, также как и упомянутые выше тригонометрические тождества.)

Построение множества Мандельброта

Теперь мы можем рассмотреть, как определяется множество Мандельброта. Пусть z — это некоторое произвольное комплексное число. Каковым бы ни было это число, оно представляется некоторой точкой на плоскости Аргана. Рассмотрим теперь *отображение*, при котором z превращается в новое комплексное число, равное

$$z \to z^2 + c,$$

где c есть некое фиксированное (то есть заданное) комплексное число. Числу $z^2 + c$ будет сопоставляться некоторая другая точка на плоскости Аргана. Например, если c равно числу $1{,}63 - i4{,}2$, то z отображается согласно формуле

$$z \to z^2 + 1{,}63 - i4{,}2,$$

так что, в частности, число 3 превратится в $3^2 + 1{,}63 - i4{,}2 = 9 + 1{,}63 - i4{,}2 = 10{,}63 - i4{,}2$,

а число $-2{,}7 + i0{,}3$ в

$$(-2{,}7 + i0{,}3)^2 + 1{,}63 - i4{,}2 = \\ = (-2{,}7)^2 - (0{,}3)^2 + 1{,}63 + \\ + i\{(-2{,}7)(0{,}3) - 4{,}2\} = 8{,}83 - i5{,}82.$$

Когда числа становятся громоздкими, вычисления лучше выполнять на компьютере.

Теперь, каково бы ни было число c, число 0 превращается, согласно принятой схеме, в число c. А что же можно сказать о самом числе c? Оно превращается в $c^2 + c$. Давайте продолжим этот процесс, применив наше преобразование к $c^2 + c$. Мы получим:

$$(c^2 + c)^2 + c = c^4 + 2c^3 + c^2 + c.$$

Снова повторим отображение, применив его к приведенному выше числу. Мы получим:

$$(c^4 + 2c^3 + c^2 + c)^2 + c = \\ = c^8 + 4c^7 + 6c^6 + 6c^5 + 5c^4 + 2c^3 + c^2 + c.$$

Потом еще раз применим процедуру, теперь уже к последнему числу, и т. д. В результате мы получаем последовательность комплексных чисел, которая начинается с числа 0:

$$0,\ c,\ c^2 + c,\ c^4 + 2c^3 + c^2 + c,\ \ldots .$$

Данная процедура, будучи реализована при *некоторых* определенных значениях комплексного числа c, дает последовательность чисел, которые все время остаются вблизи начала координат плоскости Аргана; точнее, для выбранных таким образом значений c получаемая последовательность оказывается *ограниченной*, то есть любой ее член находится в пределах некоторого фиксированного круга с центром в начале координат (рис. 3.12). Хорошим примером здесь может служить последовательность $c = 0$, поскольку каждый ее член равен 0. Другим примером ограниченного поведения является случай $c = -1$, при котором получается последовательность $0, -1, 0, -1, 0, -1, \ldots$; еще один пример — это $c = i$, когда получается последовательность $0, i, i - 1, -i, i - 1, -i, i - 1, -i, \ldots$. Однако, для целого ряда других комплексных чисел c получаемая последовательность все дальше удаляется от начала координат, то есть является *неограниченной* и не может находиться целиком в пределах фиксированного круга. Именно так происходит при $c = 1$, когда получается последовательность $0, 1, 2, 5, 26, 677, 458\,330, \ldots$; аналогичное поведение имеет место в случае $c = 3$ — соответствующая последовательность имеет вид $0, -3, 6, 33, 1086, \ldots$; а также случай $c = i - 1$, который приводит к последовательности $0,\ i - 1,\ -i - 1,\ -1 + 3i,\ -9 - i5,\ 55 + i91,\ -5257 + i10011,\ \ldots$.

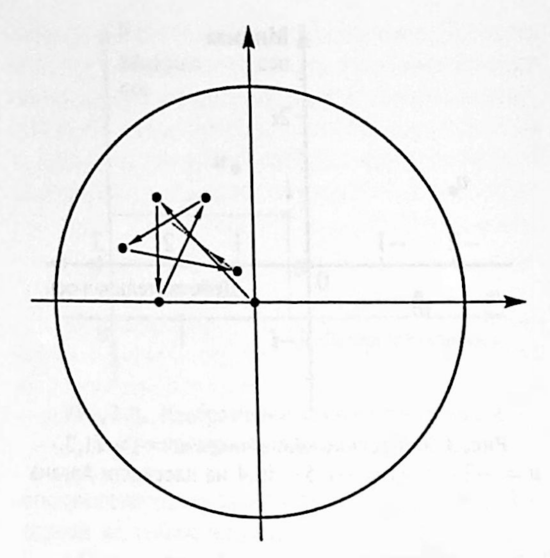


Рис. 3.12. Последовательность точек на плоскости Аргана *ограничена*, если вся она целиком помещается в пределах некоторого фиксированного круга. (Итерации на рисунке начинаются с точки 0 и построены для $c = -1/2 + (1/2)i$.)

Множество Мандельброта — то есть *зачерненная* часть страны Тор'Блед-Нам[5] — как раз и есть та самая область на плоскости Аргана, что состоит из всех точек c, для которых получаемая последовательность является ограниченной. *Белая* же область состоит из тех точек c, для которых получается неограниченная последовательность. Приведенные выше подробные рисунки основаны на результатах компьютерных вычислений. На компьютере был проведен систематический перебор всевозможных комплексных чисел c, для каждого из них строилась последовательность $0,\ c,\ c^2+c,\ \ldots$, после чего согласно некоторому критерию определялось, ограничена или нет получаемая последовательность. Если последовательность оказывалась *ограниченной*, то соответствующая

[5] В оригинале — *Tor'Bled-Nam*. А что получится, если прочитать наоборот? — *Прим. ред.*

числу c точка экрана становилась черной. Таким образом, для каждой точки в рассматриваемой области компьютер решал, закрасить ее в белый или черный цвет.

Множество Мандельброта впечатляет своей сложностью, особенно учитывая, как это часто бывает в математике, удивительную простоту его определения. Кроме того, структура этого множества в целом не очень чувствительна к выбору алгебраической формы отображения — $z \to z^2 + c$. Многие другие итеративные отображения (например, $z \to z^3+iz^2+c$) приводят к поразительно похожим структурам (при условии выбора подходящего начального числа — возможно, это не 0, а значение, четко задаваемое вполне определенным математическим правилом для каждого разумно выбранного отображения). Подобные «мандельбротовы» структуры характеризуются некоторыми универсальными или абсолютными свойствами по отношению к итеративным комплексным отображениям. Изучение таких структур является предметом отдельного раздела математики — так называемой *теории комплексных динамических систем.*

Платоническая реальность математических понятий?

Насколько реальны объекты математического мира? Некоторые считают, что ничего реального в них быть не может. Математические объекты суть просто понятия, они представляют собой мысленные идеализации, созданные математиками — часто под влиянием внешних проявлений и кажущегося порядка окружающего нас мира; но при этом они — всего лишь рожденные разумом абстракции. Могут ли они представлять собой что-либо, кроме просто произвольных конструкций, порожденных человеческим мышлением? И в то же время эти математические понятия часто выглядят глубоко реальными и эта реальность выходит далеко за пределы мыслительных процессов любого конкретного математика. Тут как будто имеет место обратное явление — человеческое мышление как бы само оказывается направляемым к некой внешней истине — истине, которая реальна сама по себе, и которая открывается каждому из нас лишь частично.

Множество Мандельброта представляет собой потрясающий пример. Его удивительно сложная структура не является результатом изобретения ни какой-либо отдельной личности, ни группы математиков. Сам Бенуа Мандельброт — американский математик польского происхождения (и один из главных разработчиков теории фракталов), который первый [3] изучил это множество, не мог себе представить, насколько фантастически сложным окажется этот объект, хотя и понимал, что обнаружил нечто очень интересное. Действительно, увидев самые первые компьютерные изображения, он счел увиденные им размытые структуры результатом сбоя (Мандельброт [1986])! И только потом он убедился, что они действительно являлись частью множества. Более того, сложную структуру множества Мандельброта во всех ее деталях не под силу охватить никому из нас, и ее невозможно полностью отобразить на компьютере. Создается впечатление, что рассматриваемая структура не является всего лишь частью нашего мышления, но что она реальна сама по себе. Кто бы из математиков или программистов ни занялся изучением этого множества, результатом их исследований обязательно будут приближения к *одной и той же единой для всех* фундаментальной математической структуре. Не важно, на каком компьютере проводятся вычисления — лишь бы он правильно работал (конечно, если отвлечься от различий в степени подробности выявляемых деталей и скорости их вывода, связанными с различиями в производительности, объеме памяти и параметрах монитора). При этом компьютер применяется в сущности так же, как прибор в руках физика-экспериментатора, исследующего строение физического мира. Множество Мандельброта — это не плод человеческого воображения, а открытие. Подобно горе Эверест, множество Мандельброта просто-напросто уже существовало «там вовне»!

Аналогичным образом сама система комплексных чисел обладает глубокой и вневременной реальностью, выходящей далеко за пределы мысленных конструкций, созданных любым конкретным математи-

ком. Первые шаги на пути к пониманию комплексных чисел связаны с работами Джероламо Кардано. Он родился и жил в Италии с 1501 по 1576 год — врач, игрок и составитель гороскопов (однажды он даже составил гороскоп для Иисуса Христа), написавший в 1545 году очень важный и оказавший большое влияние на последующее развитие математики трактат по алгебре под названием *Ars Magna*. В этом трактате он предложил первое полное решение (в терминах иррациональных выражений, то есть корней n-й степени) кубического уравнения в общем виде[6]. Кардано заметил, что в некоторых — так называемых «неприводимых» — случаях, когда уравнение имело три действительных решения, он был вынужден на определенном этапе включать в свою формулу *квадратный корень из отрицательного числа*. Хотя это обстоятельство и приводило его в замешательство, он понял, что полное решение можно получить *тогда и только тогда*, если допустить возможность *извлечения* таких квадратных корней (окончательный результат всегда оказывался действительным числом). Позднее, в 1572 году Рафаэль Бомбелли в своей работе, озаглавленной «Алгебра», обобщил работу Кардано, положив начало изучению алгебры комплексных чисел.

Хотя вначале может показаться, что введение таких квадратных корней из отрицательных чисел представляет собой всего лишь некоторый прием — математическое изобретение для достижения конкретной цели, — впоследствии становится очевидным, что потенциал этих объектов выходит далеко за рамки их использования для первоначально поставленных целей. При том, что изначально комплексные числа вводились (как уже упоминалось выше) для обеспечения возможности «безнаказанно» извлекать квадратные корни из отрицательных чисел, сделав этот шаг, мы получили в качестве бесплатного приложения еще и способ извлечения корней любой степени, а также решения любых алгебраических уравнений. Далее мы обнаружим у комплексных чисел много других волшебных свойств, о которых мы вначале даже и не подозревали. Эти свойства просто-напросто *уже существуют* «там вовне». Они не были привнесены туда ни Кардано, ни Бомбелли, ни Уоллисом, ни Котсом, ни Эйлером, ни Весселем, ни Гауссом, несмотря на несомненную прозорливость и их, и других великих математиков. Этот набор волшебных свойств был изначально присущ самой структуре, которую они шаг за шагом открывали. Когда Кардано вводил комплексные числа, он и подозревать не мог о существовании множества открытых впоследствии чудесных свойств, названных именами знаменитых ученых — таких как интегральная формула Коши, теорема отображения Римана или свойство продолжения Леви. Эти и многие другие замечательные свойства присущи самим числам — в точности тем самым числам, с которыми Кардано впервые столкнулся в 1539 году.

Что такое математика — изобретение или открытие? Процесс получения математиками результатов — что это: всего лишь построение не существующих в действительности сложных мысленных конструкций, мощь и элегантность которых способна обмануть даже их собственных изобретателей, заставив их поверить в «реальность» этих не более чем умозрительных построений? Или же математики действительно открывают истины уже где-то существующие, чья реальность в значительной степени независима от их деятельности? Я думаю, что читателю должно стать уже совершенно ясно, что я склонен придерживаться скорее второй, чем первой точки зрения, по крайней мере, в отношении таких структур, как комплексные числа или множество Мандельброта.

Однако, не все так просто. Как я уже сказал, в математике существуют вещи, к которым термин «открытие» подходит больше, чем «изобретение» — как в только что упомянутых примерах. Это происходит, когда структура дает гораздо больше того, что в нее было вложено изначально. Можно встать и на такую точку зрения, согласно которой в этих случаях математики просто наталкиваются на «творения Бога». Встречаются, однако, другие ситуации, когда математические структуры не столь убедительно уникальны — например, когда посреди доказательства какого-нибудь результата возникает необходимость в некой хитроумной,

---

[6] Частично основанное на более ранних работах Сципионе дель Ферро и Тартальи.

хотя и далеко не уникальной конструкции для достижения весьма специфической цели. В этих случаях от вновь созданной конструкции вряд ли следует ожидать больше того, что было в нее первоначально заложено, и термин «изобретение» представляется более подходящим, чем «открытие». Они, действительно, суть просто «творения человека». Согласно этой точке зрения, истинные математические открытия должны, как правило, рассматриваться как достижения более великие, чем «просто» изобретения.

Такого рода ранжирование обнаруживает некоторое сходство с тем, что мы иногда наблюдаем в области искусства или техники. Великие произведения искусства действительно «ближе к Богу», чем менее значительные творения. У художников нередко возникает чувство, что в своих величайших произведениях они открывают вечные истины, существовавшие уже до них в некотором высшем смысле [7], в то время как менее значительные произведения могут быть более случайными, являясь по своей природе всего лишь порождениями простых смертных. Точно также и новое инженерное решение с очень красивой структурой, позволяющее достичь значительных результатов через применение простой и неожиданной идеи, может с полным на то основанием рассматриваться скорее не как изобретение, а как открытие.

Однако, высказав все эти соображения, я не могу отделаться от ощущения, что в случае математики вера в некоторое высшее вечное существование — по крайней мере для наиболее глубоких математических концепций, — имеет под собой гораздо больше оснований, чем в других областях человеческой деятельности. Несомненная уникальность и универсальность такого рода математических идей по своей природе существенно отличается от всего того, с чем приходится сталкиваться в области искусства и техники. Точка зрения, согласно которой математические понятия могут существовать в такого рода вневременном, высшем смысле, была впервые высказана еще в древности (около 360 года до н. э.) великим греческим философом Платоном, и поэтому ее часто называют *математическим платонизмом*. Она играет важную роль в дальнейшем изложении.

В главе 1 я довольно много места уделил обсуждению точки зрения *сильного искусственного интеллекта*, согласно которой мыслительные явления находят свое воплощение в рамках математического понятия алгоритма. В главе 2 я особо подчеркнул, что алгоритм есть действительно очень глубокое и «Богом данное» понятие. В этой главе я старался доказать, что такие «Богом данные» математические идеи существуют в определенном смысле вне времени и независимо от нас смертных. Не могут ли эти соображения служить своего рода подтверждением справедливости концепции сильного искусственного интеллекта, допуская возможность некоего высшего существования мыслительной деятельности? Это вполне возможно — и я даже собираюсь далее привести ряд соображений в поддержку в чем-то похожей точки зрения. Но если у мыслительных явлений и вправду имеется такое вместилище, я все же не думаю, что это может относиться и к понятию алгоритма. Тут нужно что-то более «тонкое». Последующее обсуждение будет в значительной степени опираться на тот факт, что связанные с понятием алгоритма объекты составляют очень узкую и ограниченную часть математики. Следующая глава даст некоторое представление об огромных возможностях и изяществе неалгоритмической математики.

Примечания

1. См. Мандельброт [1986]. Выбранная мною конкретная последовательность коэффициентов увеличения взята из работы Пайтгена и Рихтера [1986], в которой можно познакомиться с большим количеством цветных изображений множества Мандельброта. Другие поразительные иллюстрации можно найти в книге Пайтгена и Заупе [1988].

---

[7] Как сказал выдающийся аргентинский писатель Хорхе Луис Борхес: «...знаменитый поэт в большей степени первооткрыватель, чем изобретатель...».

2. Насколько мне известно, точка зрения, согласно которой для любого действительно числа должно существовать некое — пусть неэффективное и даже совершенно неопределимое в рамках заданной формальной системы (см. главу 4) — правило, позволяющее определить его n-й знак, является вполне непротиворечивой, хотя и нетрадиционной. Я сильно надеюсь на то, что этот подход *действительно* непротиворечив, поскольку именно этой точки зрения я сам больше всего хотел бы придерживаться!

3. Первенство обнаружения этого множества до сих пор остается предметом споров (см. Брукс, Мательски [1981], Мандельброт [1989]), но сама возможность таких споров представляет собой дополнительное свидетельство в пользу того, что здесь мы имеем дело скорее с открытием, чем с изобретением.

Глава 4

Истина, доказательство и интуиция

Программа Гильберта для математики

Что есть истина? Как мы составляем наши суждения о том, что в мире является справедливым, верным, а что — нет? Следуем ли мы некоторому алгоритму, которому отдается предпочтение среди прочих, менее эффективных, в процессе всемогущего естественного отбора? Или же возможен некий иной путь — не алгоритмизированный, а основанный на особой проницательности, интуитивный, инстинктивный — позволяющий *угадывать* правду? Это представляется нелегким вопросом. Наши суждения зависят от сложных взаимосвязанных комбинаций данных, поставляемых органами чувств, и наших размышлений и догадок. Более того, во многих реальных ситуациях не может существовать единого мнения по поводу того, что *на самом деле* истинно, а что — ложно. Чтобы упростить задачу, рассмотрим только лишь *математическую* истину. Как мы формируем суждения — а может, даже и наши «стопроцентно верные» знания — при ответе на вопросы из области математики? Там уж, по крайней мере, все должно быть не так размыто, очерчено более ясно. Там не может возникать вопросов об истинности — или все-таки может? Что же, в конце концов, есть математическая истина?

Вопрос об этой истине возник не сегодня, он уходит корнями в античность, к греческим философам и математикам — и, несомненно, еще дальше, в глубь веков. Однако, несколько великих открытий и поразительных прозрений здесь были сделаны не далее как в XX столетии. Эти новые достижения заслуживают того, чтобы постараться их понять. Они носят фундаментальный характер и непосредственно касаются вопроса о том, являются ли наши мыслительные процессы полностью алгоритмизированными по своей природе или нет. Четко разобраться в этом — задача, имеющая для нас весьма важное значение.

В последней части XIX века математика шагнула далеко вперед в результате развития все более и более мощных методов математического доказательства. (Давид Гильберт и Георг Кантор, с которыми мы познакомились ранее, и великий французский математик Анри Пуанкаре, с которым нам еще предстоит встретиться, шли во главе этих разработок.) Как следствие, математики стали обретать уверенность в том, что применение этих методов приведет к успеху. Многие из таких методов основаны на рассмотрении *множеств*[1] с бесконечным числом членов, и доказательства часто оказывались осуществимы благодаря именно тому, что такое множество можно было рассматривать как

[1] «Множество» означает набор предметов — физических объектов или математических абстракций, — который может рассматриваться как единое целое. В математике элементы (т. е. члены) множества часто сами являются множествами, поскольку множества могут собираться таким образом, чтобы самим формировать множества. Тем самым можно рассматривать множества множеств, множества множеств множеств и т. д.

реальный «объект» — завершенное единое целое, существующее не только в абстракции. Многие из этих идей родились из в высшей степени оригинальной концепции Кантора о *бесконечных числах*, которую он развил, последовательно используя бесконечные множества. (Мы кратко ознакомились с ними в предыдущей главе.)

Однако эта уверенность пошатнулась, когда в 1902 году английский логик и философ Бертран Рассел придумал свой знаменитый парадокс (который предвидел и сам Кантор и который выводился непосредственно из его диагонального процесса). Чтобы понять доводы Рассела, мы сначала должны хотя бы немного почувствовать, как можно представить множество в виде единого целого. Давайте представим себе множество, характеризуемое некоторым (общим) *свойством*. Например, набор *красных* предметов может быть охарактеризован словом «краснота» как его определяющим свойством: нечто принадлежит этому множеству тогда и только тогда, когда это обладает «краснотой» (имеет красный цвет). Это позволит нам «перевернуть» точку зрения и трактовать свойство как единичный объект, который будет состоять из всего множества вещей, обладающих данным свойством. При таком рассмотрении «краснота» *эквивалентна* множеству всех красных предметов. (При этом мы можем предполагать существование «там вовне» и других множеств, члены которых не могут быть охарактеризованы подобным простым свойством.)

Идея формулировки понятий в терминах множеств послужила основой для процедуры, предложенной в 1884 году влиятельным немецким логиком Готтлибом Фреге, которая позволяла определять *числа* через множества. К примеру, что мы понимаем под числом 3? Мы знаем, в чем заключается «тройственность», но что есть число 3 само по себе? Очевидно, что «тройственность» есть свойство *наборов* объектов, т. е. свойство множеств: некоторое множество обладает данным свойством тогда и только тогда, когда это множество состоит из трех членов. Этим свойством характеризуется, скажем, тройка призеров-медалистов некоторой Олимпиады. Равно как и набор шин к трехколесному велосипеду, или листья на одном стебельке обычного клевера, или множество всех решений уравнения $x^3 - 6x^2 + 11x - 6 = 0$. Как же можно тогда определить по Фреге само число 3? Согласно Фреге, 3 — это множество множеств, а именно, *всех* множеств, имеющих свойство «тройственности» [1]. Таким образом, множество содержит три члена тогда и только тогда, когда оно принадлежит множеству 3 по Фреге.

Может показаться, что мы попадаем в замкнутый круг, но в действительности это совсем не так. Мы можем определить числа в общем случае как совокупности всевозможных эквивалентных множеств, где говоря «эквивалентные», мы понимаем «состоящие из элементов, которые могут быть попарно сопоставлены друг другу» (или, в более привычной терминологии, «имеющих одинаковое число элементов»). Тогда число 3 будет одной из этих совокупностей множеств, которая содержит в себе в качестве члена множество, состоящее, скажем, из яблока, апельсина и груши. Обратите внимание, что это принципиально отличается от определения «$\mathbb{3}$», данного Черчем (с. 89). Существуют также и другие определения, причем более популярные в наши дни.

Вернемся теперь к парадоксу Рассела. В чем он заключается? В нем рассматривается множество R, определенное следующим образом:

> R есть множество множеств, которые не являются членами самих себя.

Таким образом, R есть набор множеств X, отвечающих следующему условию: среди членов множества X не должно быть самого X.

Не является ли абсурдным предполагать, что множество в действительности может быть членом самого себя? Ничуть. Рассмотрим, к примеру, множество I, состоящее из *бесконечных* множеств (множеств с бесконечным числом членов). С очевидностью, существует бесконечное число *различных* бесконечных множеств, и само множество I, таким образом, является бесконечным. И, таким образом, оно, действительно, принадлежит самому себе! Но как же, в таком случае, рассуждения Рассела дают нам парадоксальное утверждение? Давайте спросим: является ли множество Рассела R

членом самого себя или нет? Если *нет*, то оно должно принадлежать себе, ибо R состоит как раз из таких множеств, которые не являются членами самих себя. То есть, в конечном счете, R принадлежит R — противоречие! С другой стороны, если R *есть* член самого себя, то, поскольку «самое себя» — это R, оно в то же время принадлежит множеству, члены которого, по определению, не могут быть составляющими самих себя, т. е. все-таки не принадлежит самому себе — и вновь противоречие![2)]

Этот парадоксальный вывод не был праздной игрой ума: Рассел использовал — хотя и в крайней форме — тот же тип весьма общих теоретико-множественных методов, которые математики начинали использовать в то время для своих доказательств. Становилось очевидным, что казавшаяся незыблемой почва ускользает из-под ног, и поэтому необходимо было как можно точнее определить, какие рассуждения считать допустимыми. Ясно было, что такие рассуждения должны быть свободны от внутренних противоречий, и что утверждения, которые будут выводиться с их помощью как следствия из априори верных посылок, должны быть также верными. Рассел, совместно со своим коллегой Альфредом Нортом Уайтхедом, взялся за развитие такой полностью формализованной системы аксиом и правил вывода, на язык которой стало бы возможным перевести все виды корректных математических рассуждений. Все правила подвергались тщательному отбору, дабы избежать «ложных» путей рассуждений, могущих привести к парадоксам, подобным упомянутому выше. Однако схема, появившаяся на свет в результате этих усилий, была очень громоздка и оказалась весьма ограниченной по диапазону различных типов математических рассуждений, которые она охватывала. Великий математик Давид Гильберт (которого мы впервые встретили в главе 2) задался целью создать более практичную и универсальную систему. В нее должны были войти *все* типы математических рассуждений из *всех* областей математики. Более того, Гильберт стремился сделать возможным *строгое доказательство* отсутствия противоречий в своей схеме. Тогда математика раз и навсегда смогла бы встать на прочную и неколебимую основу.

Однако надежды Гильберта и его последователей были перечеркнуты, когда в 1931 году блестящий австрийский логик и математик Курт Гедель выдвинул поразительную теорему, которая до основания разрушала программу Гильберта. Гедель показал, что *любая* подобная точная («формальная») система аксиом и правил вывода, если только она достаточна широка, чтобы содержать в себе описания простых арифметических теорем (как, например, «последняя теорема Ферма», рассмотренная в главе 2), и если она свободна от противоречий — то такая система должна включать утверждения, которые не являются ни доказуемыми, ни недоказуемыми в рамках формализма данной системы. Истинность таких «неразрешимых» утверждений, следовательно, не может быть выяснена с помощью методов, допускаемых самой системой. Более того, Гедель смог показать, что даже утверждение о непротиворечивости системы аксиом, будучи переведенным в форму соответствующей теоремы, само по себе является «неразрешимым». Для нас будет очень важным понять природу этой неразрешимости. Тогда мы увидим, почему выводы Геделя опровергали самое основание программы Гильберта. Мы также увидим, каким образом они дают нам возможность, воспользовавшись интуицией, выходить за пределы любой рассматриваемой формализованной математической системы. Это понимание будет решающим для того, чтобы, в свою очередь, лучше понять обсуждаемое далее.

Формальные математические системы

Необходимо будет несколько уточнить, что мы понимаем под «формальными математическими системами аксиом и правил вывода». Мы должны предположить на-

[2)] Можно дать занятную трактовку парадокса Рассела в более привычных терминах. Представьте себе библиотеку с двумя каталогами, один из которых перечисляет только те книги в библиотеке, которые хотя бы раз ссылаются на себя самих, а другой — остальные книги, т. е. те, которые не упоминают себя. В каком из этих каталогов, в таком случае, должен фигурировать второй каталог?

личие некоторого алфавита символов, через которые будут записываться математические выражения. Эти символы в обязательном порядке должны быть адекватны для записи натуральных чисел с тем, чтобы в нашу систему могла быть включена «арифметика». По желанию, мы можем использовать общепринятую арабскую запись 0, 1, 2, 3, ..., 9, 10, 11, 12, ..., хотя при этом конкретные выражения для правил вывода становятся несколько более сложными, чем требуется. Гораздо более простые выражения получаются, скажем, при использовании записи вида 0, 01, 011, 0111, 01111, ... для обозначения последовательности натуральных чисел (или, в качестве компромисса, мы могли бы использовать двоичную запись). Однако, поскольку это могло бы стать источником разночтений в дальнейших рассуждениях, я буду для простоты придерживаться обычной арабской записи независимо от способа обозначения, которая может *на самом деле* использоваться в данной системе. Нам мог бы понадобиться символ «пробел» для разделения различных «слов» или «чисел» в нашей системе, но, так как это тоже может вызвать путаницу, то мы будем по мере необходимости использовать для этих целей просто запятую (,). Произвольные («переменные») натуральные числа (равно как и целые, рациональные и т. д.; но давайте здесь ограничимся натуральными) мы станем обозначать буквами, например, $t, u, v, w, x, y, z, t', t'', t'''$ и т. п. Штрихованные буквы $t', t'', \ldots$ вводятся нами в употребление, дабы не ограничивать число переменных, которые могут встретиться в произвольном выражении. Мы будем считать *штрих* ($'$) отдельным символом формальной системы, так что действительное количество *символов* в системе остается конечным. Помимо этого нам также потребуются символы для базовых арифметических операций $=, +, \times$ и т. д.; для различных видов скобок (,), [,], и для обозначения логических операций, таких как & («и»), $\Rightarrow$ («следует»), $\vee$ («или»), $\Leftrightarrow$ («тогда и только тогда»), $\sim$ («не»). Дополнительно нам будут нужны еще и логические «кванторы»: квантор существования $\exists$ («существует... такое, что») и квантор общности $\forall$ («для любого... выполняется»). Тогда мы сможем такие утверждения, как, например, «последняя теорема Ферма», привести к виду:

$$\sim \exists\, w, x, y, z\big[(x+1)^{w+3} + (y+1)^{w+3} = \\ = (z+1)^{w+3}\big]$$

(см. главу 2, с. 82). (Я мог бы написать «0111» для «3», и, возможно, использовать для «возведения в степень» обозначение, более подходящее к рассматриваемому формализму; но, как уже говорилось, я буду придерживаться стандартной системы записи во избежании ненужной путаницы.) Это утверждение (если читать его до левой квадратной скобки) звучит как:

> «Не существует таких натуральных чисел w, x, y, z, что...».

Мы можем также переписать последнюю теорему Ферма при помощи $\forall$:

$$\forall\, w, x, y, z\big[\sim (x+1)^{w+3} + (y+1)^{w+3} = \\ = (z+1)^{w+3}\big],$$

которое будет читаться следующим образом (заканчивая символом «не» после левой квадратной скобки):

> «Для любых натуральных чисел w, x, y, z не может быть выполнено...»,

что логически эквивалентно написанному ранее.

Нам понадобятся еще и буквы, обозначающие целые утверждения, для чего я буду использовать заглавные буквы P, Q, R, S, ... Таким утверждением может, к примеру, служить и вышеприведенная теорема Ферма:

$$F = \sim \exists\, w, x, y, z\big[(x+1)^{w+3} + (y+1)^{w+3} = \\ = (z+1)^{w+3}\big].$$

Утверждение может также зависеть от одной или более переменных; например, нас может интересовать формулировка теоремы Ферма для некоторого конкретного[3] значения

[3] Хотя справедливость теоремы Ферма в общем случае пока не доказана, ее справедливость для некоторых частных случаев, таких как $G(0)$, $G(1)$, $G(2)$, $G(3)$, доказана вплоть до $G(125\,000)$. Другими словами, доказано, что куб никакого числа не может быть суммой кубов двух других положительных чисел, четвертая степень числа не может быть суммой четвертых степеней других чисел и т. д. вплоть до степени 125 000. (Несколько лет назад теорема Ферма была доказана в общем виде. См. сноску 5 на с. 82. — *Прим. ред.*)

степени $w+3$

$$G(w) = \sim \exists\, x, y, z\left[(x+1)^{w+3} + (y+1)^{w+3} = (z+1)^{w+3}\right],$$

так что $G(0)$ утверждает, что «куб не может быть суммой кубов положительных чисел»; $G(1)$ говорит о том же применительно к четвертым степеням и так далее. (Обратите внимание на отсутствие w после символа $\exists$.) Тогда теорема Ферма гласит, что $G(w)$ выполняется для любого w:

$$F = \forall\, w[G(w)].$$

$G(\ )$ является примером так называемой функции исчисления высказываний, т. е. утверждением, которое зависит от одной или более переменных.

Аксиомы нашей системы будут представлять из себя перечень утверждений общего характера, чья справедливость в рамках принятого символизма предполагается самоочевидной. Например, для произвольных утверждений или *функций исчисления высказываний* P, Q, $R(\ )$ мы могли бы указать среди прочих аксиом системы такие, как

$$(P\&Q) \Rightarrow P,$$
$$\sim(\sim P) \Leftrightarrow P,$$
$$\sim \exists\, x[R(x)] \Leftrightarrow \forall\, x[\sim R(x)],$$

«априорная истинность» которых уже заключена в их смысловых значениях. (Первое утверждение означает лишь, что «если выполняется P и Q, то выполняется и P»; второе устанавливает равносильность утверждений «неверно, что не выполняется P» и «P выполняется»; а третье может быть проиллюстрировано эквивалентностью двух способов формулировки теоремы Ферма, данных выше.) Мы можем также включить основные аксиомы арифметики:

$$\forall\, x, y[x+y = y+x],$$
$$\forall\, x, y, z[(x+y)\times z = (x\times z) + (y\times z)],$$

хотя некоторые предпочитают определять арифметические операции через более простые понятия и выводить вышеуказанные утверждения как теоремы. *Правила вывода* могут вводиться в виде (самоочевидных) процедур типа

«Из P и $P \Rightarrow Q$ следует Q».

«Из $\forall\, x[R(x)]$ мы можем вывести любое утверждение, получающееся путем подстановки конкретного натурального числа x в $R(x)$».

Такие правила являются инструкциями, следуя которым, можно с помощью утверждений, чья истинность уже доказана, получать новые утверждения.

Теперь, отталкиваясь от системы аксиом и раз за разом применяя правила вывода, мы имеем возможность построить достаточно длинные цепочки новых утверждений. На любой стадии этого процесса мы можем использовать снова и снова любую из аксиом, а также обратиться к любому из уже выведенных нами производных утверждений. Каждое утверждение из корректно выстроенной цепочки называется теоремой (несмотря на то, что многие из них достаточно тривиальны и неинтересны с точки зрения математики). Если у нас есть некое утверждение P, которое мы хотим *доказать*, то мы должны подобрать такую цепочку, выстроенную в согласии с действующими правилами вывода, которая заканчивается утверждением P. Такая цепочка предоставит нам доказательство P в рамках системы; а P тогда будет являться, соответственно, теоремой.

Идея программы Гильберта состояла в том, чтобы найти применительно к любой отдельно взятой области математики набор аксиом и правил вывода, который был бы достаточно полным для *всех* возможных в данной области корректных математических рассуждений. Пусть такой областью будет арифметика (с добавленными кванторами $\exists$ и $\forall$, позволяющими формулировать утверждения, подобные последней теореме Ферма). То, что мы не рассматриваем более общую область математики, не умаляет нашу задачу: арифметика и сама по себе обладает общностью, достаточной для применения процедуры Геделя. Если мы допустим, что благодаря программе Гильберта мы действительно располагаем такой всеобъемлющей системой аксиом и правил вывода для арифметики, то мы тем самым обретаем и определенный критерий для выявления «корректности» математического доказательства любого утверждения в области арифметики. Возлагались надежды на то, что подобная

система аксиом и правил может быть *полной* в смысле предоставляемой нам принципиальной возможности решать, истинно или ложно *произвольное* утверждение, сформулированное в рамках этой системы.

Гильберт рассчитывал, что для любой строки символов, представляющих математическое утверждение, скажем, P, можно будет доказать либо P, либо $\sim P$, если P истинно или ложно, соответственно. Здесь мы в обязательном порядке оговариваем, что строка должна быть *синтаксически корректна*, где «синтаксически корректна» по сути означает «грамматически корректна» — то есть удовлетворяет всем правилам записи, принятым в данном формализме, среди которых будет правильное попарное соответствие скобок и т. п. — так чтобы P всегда имело четко определенное значение «ложь» или «истина». Если бы надежды Гильберта оправдались, то можно было бы вообще не задумываться о том, что означает то или иное утверждение! P было бы просто-напросто синтаксически корректной строкой символов. Строке было бы приписано значение **ИСТИНА**, если бы P являлось теоремой (другими словами, если бы P было доказуемо в рамках системы); или же **ЛОЖЬ**, если бы теоремой было $\sim P$. Чтобы такой подход имел смысл, мы должны дополнительно к условию полноты наложить еще и условие *непротиворечивости*, гарантирующее отсутствие такой строки символов P, для которой как P, так и $\sim P$ были бы теоремами. Ведь в противном случае P могло бы быть одновременно и **ИСТИНОЙ**, и **ЛОЖЬЮ**!

Такой подход, согласно которому можно пренебрегать смысловыми значениями математических выражений и рассматривать их лишь как строки символов некоторой формальной математической системы, в математике получил название *формализма*. Некоторым нравится эта точка зрения, с которой математика превращается в своего рода «бессмысленную игру». Однако я сам не являюсь сторонником таких идей. Все-таки именно «смысл» — а не слепые алгоритмические вычисления — составляет сущность математики. К счастью, Гедель нанес формализму сокрушающий удар! Давайте посмотрим, как он это сделал.

Теорема Геделя

Часть доказательства, приведенного Геделем, содержало некий очень сложный и детализированный кусок. Однако нам не обязательно разбираться во всех его тонкостях. Основная идея, в то же время, была проста, красива и глубока. И ее мы сможем оценить по достоинству. В «сложной» части (которая, впрочем, содержит много остроумных рассуждений) подробно показано, каким образом частные правила вывода и использование различных аксиом формальной процедуры могут быть представлены в виде *арифметических операций*. (Хотя в сложной части становится понятной плодотворность этих действий!) Для этого представления нам необходимо будет найти какой-нибудь удобный способ нумерации утверждений при помощи натуральных чисел. Один из способов мог бы заключаться в том, чтобы использовать своего рода «алфавитный» порядок для строчек символов формальной системы, имеющих одинаковую длину, упорядочить заранее строчки по длине. (Таким образом, за выстроенными в алфавитном порядке строками из одного символа будут следовать строки длиной в два символа, также упорядоченные по алфавиту; за ними идут строки из трех символов и так далее.) Это называется *лексикографическим порядком*[4]. В действительности Гедель использовал более сложную систему нумерации, но различия в данном случае для нас несущественны. Нас же должны в особенности интересовать *функции исчисления высказываний* одной переменной, наподобие введенной выше $G(w)$. Пусть n-я (из пронумерованных выбранным способом строк символов) такая функция от аргумента w

[4] Мы можем представить себе лексиграфический способ упорядочивания как обычный способ, используемый для натуральных чисел, только сделанный «по основанию $k+1$», где для $k+1$ чисел берутся различные символы формальной системы, вместе с новым «нулем», который никогда не используется. (Последняя сложность возникает в связи с тем, что числа, начинающиеся с нуля, и те, где он опущен — равны.) Простое лексикографическое упорядочивание в строчках из девяти символов осуществляется при помощи натуральных чисел, которые могут быть выписаны в стандартной десятичной системе без нуля: 1, 2, 3, 4, ..., 8, 9, 11, 12, ..., 19, 21, 22, ..., 99, 111, 112, ...

обозначается

$$P_n(w).$$

Мы можем допустить, чтобы наша нумерация по желанию была несколько «либеральна» в отношении синтаксически некорректных выражений. (Это позволит значительно упростить перевод системы на язык арифметических операций по сравнению со случаем, когда мы будем стараться исключить из рассмотрения синтаксически некорректные выражения.) Если $P_n(w)$ синтаксически корректно, то оно будет представлять из себя некоторое совершенно определенное арифметическое выражение, в котором фигурируют два натуральных числа n и w. Каков будет *конкретный вид* этого выражения — зависит от особенностей системы нумерации, которую мы выбрали. Но эти детали рассматриваются в «сложной» части и сейчас нас не касаются. Пусть

$$\Pi_n$$

будет n-м доказательством. (Опять же мы можем использовать «либеральную нумерацию», когда для некоторых значений n выражение Π_n не является синтаксически корректным и, тем самым, не доказывает никакую теорему.)

А теперь рассмотрим следующую *функцию исчисления высказываний* от натурального числа w:

$$\sim \exists\, x[\Pi_x \text{ доказывает } P_w(w)].$$

В выражении в квадратных скобках частично присутствуют слова, но, тем не менее, это — абсолютно точно определенное выражение. Оно говорит о том, что доказательство номер x является доказательством утверждения $P_w(\;)$, примененного к самому w. Находящийся за скобками квантор существования с отрицанием позволяет исключить из рассмотрения одну из переменных («не существует такого x, что...»), приводя нас в конечном счете к арифметической *функции исчисления высказываний*, зависящей только от w. В целом данное выражение утверждает, что *не* существует доказательства $P_w(w)$. Я буду предполагать, что оно оформлено синтаксически корректным образом (даже если $P_n(w)$ некорректно — поскольку тогда выражение было бы истинным за невозможностью существования доказательства синтаксически некорректного утверждения). На самом деле, в результате сделанного нами перевода на язык арифметики, написанное выше будет в действительности неким *арифметическим* выражением, включающим натуральное число w (тогда как в квадратных скобках окажется четко определенное арифметическое выражение, связывающее *два* натуральных числа x и w). Конечно, возможность представления этого выражения в арифметическом виде далеко не очевидна, но она существует. Рассуждения, приводящие к этому заключению, составляют наиболее трудную задачу в «сложной» части доказательства Геделя. Как и ранее, *непосредственный* вид арифметического выражения будет зависеть от способа нумерации и в еще большей степени от конкретной структуры аксиом и правил вывода, принятых в нашей системе. Поскольку все это входит в «сложную» часть доказательства, то в данном случае нас не интересует.

Мы пронумеровали все *функции исчисления высказываний*, зависящие от одной переменной, поэтому той, которую мы ввели выше, также должен быть приписан номер. Пусть этот номер будет k. Наша функция будет в таком случае k-й в общем списке. То есть

$$\sim \exists\, x[\Pi_x \text{ доказывает } P_w(w)] = P_k(w).$$

Теперь исследуем эту функцию при определенном значении: $w = k$. Мы получаем:

$$\exists\, x[\Pi_x \text{ доказывает } P_k(k)] = P_k(k).$$

Данное утверждение $P_k(k)$ является абсолютно точно определенным (синтаксически корректным) арифметическим выражением. Может ли оно быть доказано в рамках нашей формальной системы? А его отрицание $\sim P_k(k)$ — имеет ли оно такое доказательство? Ответ в обоих случаях будет отрицательный. Мы можем убедиться в этом путем исследования *смысла*, который лежит в основании процедуры Геделя. Хотя $P_k(k)$ является просто арифметическим выражением, последнее было построено нами таким образом, что написанное в левой части утверждает следующее: «внутри системы не существует доказательства $P_k(k)$». Если мы были аккуратны в определении аксиом и процедур вывода, и не ошиблись при нумерации, то тогда в рамках системы такого доказательства

найти невозможно. Если же доказательство существует, то значение утверждения, содержащегося в $P_k(k)$ — о том, что такого доказательства нет, — будет ложным, а вместе с ним будет ложным и арифметическое выражение, отвечающее $P_k(k)$. Но наша формальная система не может быть построена настолько плохо, чтобы включать в себя ложные утверждения, которые могут быть доказаны! Таким образом, в действительности, доказательство $P_k(k)$ быть не может. Но это в точности то самое, о чем говорит нам $P_k(k)$. То, что утверждает $P_k(k)$, обязано, следовательно, быть *верным*, а поэтому $P_k(k)$ должно быть верным как арифметическое выражение. Значит, мы нашли *истинное* утверждение, которое *недоказуемо в рамках системы*!

А как насчет $\sim P_k(k)$? Из предыдущих рассуждений видно, что доказательство этому утверждению внутри системы мы найти не сможем. Мы только что установили, что $\sim P_k(k)$ должно быть ложным (ибо $P_k(k)$ является истинным), а мы, по определению, не имеем возможности доказывать ложные утверждения в рамках системы! Таким образом, ни $P_k(k)$, ни $\sim P_k(k)$ недоказуемы в нашей формальной системе, что и составляет теорему Геделя.

Математическая интуиция

Обратите внимание, что мы здесь сталкиваемся с одной примечательной особенностью. Часто думают, что теорема Геделя имеет, в некотором роде, отрицательный смысл, поскольку она указывает на принципиальные ограничения в применении формальных математических рассуждений. Независимо от нашего мнения об универсальности применяемого подхода, всегда найдутся утверждения, которые не попадают в сферу его действия. Но насколько, в действительности, нас могут затрагивать частные случаи типа $P_k(k)$? В ходе предыдущих рассуждений мы установили, что $P_k(k)$ — *истинное* утверждение! Мы смогли это сделать несмотря на то, что это утверждение формально недоказуемо в рамках системы. А вот математических формалистов это *должно* волновать, потому что наши рассуждения с необходимостью приводят к выводам о неполноте их понятия «истины». Какая бы (непротиворечивая) формальная система не использовалась для арифметики, в ней будут содержаться утверждения, понимаемые нами как истинные, но которым не может быть приписано значение **ИСТИНА** при помощи вышеописанной формальной процедуры. Способ, при помощи которого формалист сумел бы обойти подобные трудности, мог бы состоять в том, чтобы не говорить о понятии истины, а только лишь о доказуемости внутри *конкретной* формальной системы. Однако же, такой подход весьма ограничен. Он не позволил бы даже сформулировать утверждение Геделя и осуществить его доказательство, как это было сделано выше, поскольку в значительной части рассуждений речь идет как раз об определении того, что есть ложь, а что — истина [2]. Некоторые формалисты встают на более «прагматическую» точку зрения, заявляя, что их не волнуют утверждения, подобные $P_k(k)$, поскольку они исключительно сложны и не интересны в качестве арифметических выражений. Отстаивают они свою точку зрения примерно так:

> «Да, есть странные утверждения, вроде $P_k(k)$, для которых мое понятие доказуемости или **ИСТИНЫ** расходится с вашим интуитивным понятием истинности, но подобные выражения едва ли встречаются в серьезной математике (по крайней мере не в такой, которая меня интересует), поскольку они абсурдно усложнены и неестественны для математики».

Несомненно, что утверждения вида $P_k(k)$, будучи полностью выписанными, были бы чрезвычайно громоздки и выглядели бы странно для числовых математических выражений. Однако за последнее время были выдвинуты сравнительно простые выражения приемлемого с точки зрения математики характера, которые эквивалентны утверждениям Геделя [3]. Они недоказуемы на основании обычных аксиом арифметики, однако же следуют из некоего свойства «самоочевидности», которым обладает сама система аксиом.

Отсутствие интереса к «математической истине», исповедуемое формалистами, кажется мне очень странной позицией в приложении к философии математики. Более

того: она совсем не так прагматична, как представляется. Когда математики проводят свои выкладки, они не намерены постоянно проверять, могут ли они быть сформулированы посредством аксиом и правил вывода некоторой сложной формальной системы. Единственно, что необходимо — быть уверенным в правомерности использования этих рассуждений для установления истины. Доказательство Геделя удовлетворяет этому требованию, так что $P_k(k)$ является математической истиной с таким же правом, как и любое другое утверждение, полученное более стандартным путем с использованием изначально заданных аксиом и правил вывода.

Процедура, которая напрашивается сама собой, заключается в следующем. Давайте положим, что $P_k(k)$ — совершенно верное утверждение (переобозначим его здесь как G_0). Тогда мы можем присоединить его к нашей системе в качестве дополнительной аксиомы. Естественно, что наша новая система будет, в свою очередь, содержать *новое* утверждение Геделя, скажем, G_1, которое также будет истинным числовым выражением. Соответственно, мы можем и G_1 добавить в нашу систему. Это даст нам новую улучшенную систему, которая также содержит новое утверждение Геделя G_2 (опять же совершенно справедливое); и мы сможем снова добавить его к системе, получая следующее утверждение Геделя G_3, которое мы тоже присоединяем — и так далее, повторяя этот процесс неограниченно. Что мы можем сказать о получившейся в результате системе, где мы используем *весь* набор $G_0, G_1, G_2, G_3, \ldots$ как дополнительные аксиомы? Может ли *эта* система быть полной? Поскольку мы теперь имеем неограниченную (бесконечную) систему аксиом, то возможность применения процедуры Геделя совсем не очевидна. Однако, это последовательное включение утверждений Геделя является в высшей степени систематичной схемой, результат применения которой может быть истолкован как обычная конечная система аксиом и правил вывода. Эта система будет иметь свое собственное утверждение Геделя G_ω, которое мы также сможем к ней присоединить, получая новую систему и с ней — еще одно утверждение Геделя $G_{\omega+1}$. Продолжая, как и ранее, мы получаем набор утверждений $G_\omega, G_{\omega+1}, G_{\omega+2}, G_{\omega+3}, \ldots$, каждое из которых истинно и может быть включено в нашу формальную систему. Сохраняя свойство строгой систематичности, этот процесс вновь приводит нас к созданию новой системы, которая охватывает все созданные к этому моменту аксиомы. Но и эта система, в свою очередь, имеет свое собственное утверждение Геделя, скажем, $G_{\omega+\omega}$ — которое можно переписать как $G_{\omega 2}$, и мы можем начать всю процедуру заново. В результате этого мы получим новый бесконечный, но систематический, набор аксиом $G_{\omega 2}, G_{\omega 2+1}, G_{\omega 2+2}$, и т. д., приводящий к еще одной новой системе — и новому утверждению Геделя $G_{\omega 3}$. Воспроизводя весь процесс, мы получаем $G_{\omega 4}$, потом — $G_{\omega 5}$ и так далее. И эта схемы также будет полностью систематичной и даст свое собственное утверждение Геделя G_{ω^2}.

Есть ли логическое завершение у этого процесса? В определенном смысле — нет; но это приводит нас к ряду трудных математических рассуждений, которые здесь не могут быть нами рассмотрены во всех деталях. Вышеуказанная процедура обсуждалась Аланом Тьюрингом в статье [4], опубликованной в 1939 году. Примечательно, что на самом деле *любое истинное* (в общепринятом смысле) утверждение в арифметике может быть получено путем повторения процедуры «геделизации» такого рода (см. Феферман [1988]). Однако это может вызвать вопрос о том, как мы в действительности решаем, является ли утверждение истинным или ложным. Исключительно важным будет также понять, как на каждом этапе нужно выполнять присоединение бесконечного семейства утверждений Геделя, чтобы они порождали единственную дополнительную аксиому (или конечное число аксиом). Для выполнения такого присоединения требуется определенная алгоритмическая систематизация нашего бесконечного семейства. Чтобы быть уверенным в том, что подобная систематизация *корректна* и приводит к желаемому результату, нам придется опереться на *интуитивные представления*, выходящие за рамки системы — точь-в-точь, как мы это сделали для установления истинности

$P_k(k)$. Именно эти «прозрения» и не могут быть систематизированы, не говоря о том, что они должны лежать вне сферы действия *любой* алгоритмической процедуры!

Интуитивная догадка, которая позволила нам установить, что утверждение Геделя $P_k(k)$ является на самом деле истинным, представляет собой разновидность общей процедуры, известной логикам как *принцип рефлексии*: посредством нее, размышляя над *смыслом* системы аксиом и правил вывода и убеждаясь в их способности приводить к математическим истинам, можно преобразовывать интуитивные представления в новые математические выражения, невыводимые из тех самых аксиом и правил вывода. То, как нами была выше установлена истинность $P_k(k)$, как раз базировалось на применении этого принципа. Другой принцип рефлексии, имеющий отношение к доказательству Геделя (хотя и не упомянутый выше), опирается на вывод новых математических истин исходя из представления о том, что система аксиом, которую мы полагаем априори адекватной для получения математических истин, является *непротиворечивой*. Применение принципов рефлексии часто подразумевает размышления о бесконечных множествах, и при этом нужно быть всегда внимательным и остерегаться рассуждений, которые могут привести к парадоксам наподобие расселовского. Принципы рефлексии полностью противопоставляются рассуждениям формалистов. Если использовать их аккуратно, то они позволяют вырваться за жесткие рамки любой формальной системы и получить новые, основанные на интуитивных догадках, представления, которые ранее казались недостижимыми. В математической литературе могло бы быть множество приемлемых результатов, чье доказательство требует «прозрений», далеко выходящих за рамки исходных правил и аксиом стандартной формальной системы арифметики. Все это свидетельствует о том, что деятельность ума, приводящая математиков к суждениям об истине, не опирается непосредственно на некоторую определенную формальную систему. Мы убедились в истинности утверждения Геделя $P_k(k)$, хотя мы и не можем вывести ее из аксиом системы. Этот тип «видения», используемый в принципе рефлексии, требует математической интуиции, которая не является результатом чисто алгоритмических операций, представимых в виде некоторой формальной математической системы. Мы вернемся к этому вопросу в главе 10.

Читатель может заметить определенное сходство между рассуждениями, устанавливающими, вопреки «недоказуемости», истинность $P_k(k)$, и парадоксом Рассела. Помимо этого, наблюдается сходство и с доказательством Тьюринга о невозможности существования «машины Тьюринга», которая могла бы решить проблему остановки. Эти сходства не случайны. Между этими тремя событиями имеется прочная историческая нить. Тьюринг пришел к своему доказательству после изучения работ Геделя. Сам Гедель был очень близко знаком с парадоксом Рассела и смог преобразовать те парадоксальные рассуждения, которые уводили слишком далеко в область логических абстракций, в состоятельное математическое доказательство. (Все эти утверждения уходят корнями к диагональному процессу Кантора, описанному в предыдущей главе, с. 100.)

Почему мы должны принимать доказательства Геделя и Тьюринга и в то же время сбрасывать со счетов рассуждения, ведущие к парадоксу Рассела? Первые являются более ясными и безупречными с точки зрения математики, тогда как парадокс Рассела строится на более туманных рассуждениях об «огромных» множествах. Но нужно признать, что различия здесь не настолько очевидны, как нам хотелось бы. Попытка придать этим различиям ясность была лейтмотивом всей идеи формализма. Доказательство Геделя, с одной стороны, показывает, что строгий формальный подход не выдерживает критики, но с другой стороны, оно не приводит нас к абсолютно надежной альтернативе. По-моему, этот вопрос до сих пор не разрешен. Процедура, используемая в современной математике с целью избежать рассуждений, вовлекающих в рассмотрение «огромные» множества и приводящих к парадоксу Расcела, не является полностью удовлетворительной[5]. Более того, она, как правило, формулируется в чисто

[5] Делается различие между «множествами» и «классами», где «множества» могут быть собраны вместе для

формалистских терминах — или же в терминах, которые не дают нам полной уверенности, что в результате их использования не возникнет противоречий.

Как бы там ни было, мне кажется, что из доказательства Геделя следует с очевидностью, что понятие математической истины не может быть заключено ни в одну из формальных систем. Математическая истина выходит за рамки любого формализма. Возможно, это ясно даже без теоремы Геделя. Иначе как бы мы решали, какие аксиомы и правила вывода брать в расчет при построении формальной системы? Нашим руководством в принятии такого решения должно всегда служить интуитивное понимание о том, что является «самоочевидно верным» с учетом «смысловых значений» символов системы. Как нам решить, какие формальные системы стоит использовать (в соответствии с нашим интуитивным ощущением «самоочевидности» и «смысла»), а какие — нет? Понятие «внутренней непротиворечивости» явно не подходит для этой цели. Можно иметь много внутренне непротиворечивых систем, которые «бессмысленны» с точки зрения их практического использования, в которых аксиомы и правила вывода имеют ложные в нашем понимании значения или же не имеют никаких. «Самоочевидность» и «смысл» — это понятия, которые потребовались бы даже без теоремы Геделя.

Однако, без этой теоремы могло бы сложиться впечатление, что интуитивные понятия «самоочевидность» и «смысл» могли бы быть использованы только в самом начале раз и навсегда, просто чтобы изначально задать формальную систему, а затем мы могли бы отказаться от них при построении строгого математического доказательства для определения истины. Тогда, в соответствии с формалистскими воззрениями, эти «расплывчатые» интуитивные понятия задействовались бы только в «предварительных» размышлениях математиков, направленных на отыскание подходящего формального доказательства; а потом, когда дело дойдет до определения математической истины, они уже не играли бы никакой роли. Теорема Геделя демонстрирует, что такой подход в действительности не является логически состоятельным в рамках фундаментальной философии математики. Понятие математической истины выходит за пределы всей теории формализма. В этом понятии есть нечто абсолютное и «данное свыше». И это как раз то, о чем трактует математический платонизм, обсуждаемый в конце предыдущей главы. Всякая формальная система имеет свойство сиюминутности и «человеко-зависимости». Такие системы, безусловно, играют очень важную роль в математических рассуждениях, но они могут указывать только частично верное (или приблизительное) направление к истине. Настоящая математическая истина выходит за пределы сотворенного человеком.

Платонизм или интуиционизм?

Я указал две противостоящие друг другу школы математической философии, решительно причисляя себя более к платонистскому, нежели к формалистскому воззрению. В действительности же я применил довольно упрощенный подход при их разделении. Существует множество тонкостей, которые можно было бы принять в расчет. Например, в рамках платонизма можно поставить вопрос о том, существуют ли в реальности объекты математической мысли или это только лишь понятие «математической истины», которое является абсолютным. Я решил не обсуждать здесь подобные различия. В моем представлении абсолютность математической истины и платонистское существование математических понятий, по существу, тождественны. «Существование», которое должно быть приписано множеству Мандельброта, к примеру, есть свойство его абсолютной природы. Принадлежит ли точка плоскости Аргана множеству

образования других множеств или классов; а классы не могут образовывать сколько-нибудь более крупные объединения, будучи для этого «слишком большими». Однако не существует правила, согласно которому можно было бы решать, какие объединения могут рассматриваться как множества, а какие с необходимостью должны быть только классами — если не считать «порочное» замкнутое правило, гласящее, что множествами являются те объединения, которые можно составлять вместе, чтобы получать новые объединения!

Мандельброта или нет — вопрос абсолютный, не зависящий от математика или компьютера, которые его исследуют. Эта «независимость-от-математика» множества Мандельброта и обеспечивает ему платонистское существование. Более того, наиболее тонкие детали этого множества лежат за пределами того, что можно достигнуть с помощью компьютера. Эти устройства способны только аппроксимировать структуры, имеющие свое, более глубокое и «не зависящее-от-компьютера», существование. Я, однако, готов согласиться с тем, что имеются и прочие разумные точки зрения, с которых можно исследовать этот вопрос. Но здесь нам нет необходимости придавать значение этим различиям.

Есть также отличие в том, насколько далеко в своем платонизме готов зайти человек, провозглашающий свою принадлежность к этой школе. Сам Гедель был глубоко убежденным платонистом. Математические выражения, которые я до сих пор рассматривал, являют собой довольно «мягкие» примеры того, что может встретиться в этом направлении [5]. Вполне возможны и более «запутанные» выражения, особенно в теории множеств. Когда рассматриваются все мыслимые ответвления этой теории, то порой возникают множества столь громадные и причудливо сконструированные, что даже такой весьма убежденный платонист, как я, может начать сомневаться в абсолютности их существования (или, напротив, несуществования) [6]. Может наступить момент, когда определения множеств становятся настолько сложными и концептуально шаткими, что вопрос об истинности или ложности относящихся к ним математических выражений становится скорее субъективным и зависящим от мнения исследователя, нежели «ниспосланным свыше». Готов ли иной математик безоглядно следовать вместе с Геделем путем платонизма, провозглашая истинность или ложность математических выражений, оперирующих подобными огромными множествами, всегда абсолютными (или «платонистскими») по своей природе; или же он, не заходя слишком далеко, будет говорить об абсолютности этих понятий лишь в том случае, если множества окажутся не слишком велики и довольно конструктивны. Ответ на этот вопрос не имеет большого отношения к нашей дискуссии. Множества (конечные или бесконечные), которые будут иметь для нас значение, по меркам вышеупомянутых множеств выглядят до смешного маленькими! Так что различия между разными платонистскими течениями нас волновать не должны.

Имеются, однако, и иные точки зрения в математике, такие как *интуиционизм* (и *финитизм*), которые, впадая в противоположную крайность, отказываются признавать существование каких бы то ни было бесконечных множеств[6)]. Интуиционизм был основан в 1924 году датским математиком Лейтзеном Э. Брауэром как альтернативный ответ — отличный от предлагаемого формализмом — на парадоксы (типа расселовского), которые могут возникать там, где бесконечные множества используются слишком вольно в математических рассуждениях. Зачатки этого подхода прослеживаются еще во времена Аристотеля, который, будучи учеником Платона, тем не менее отвергал его взгляды на абсолютное существование математических сущностей и возможность рассмотрения бесконечных множеств. Согласно интуиционизму, существование множества (бесконечного, равно как, впрочем, и конечного) не может признаваться как свойство, изначально ему присущее, а только лишь как функция правил, по которым оно организовано.

Характерная черта интуиционизма Брауэра состоит в отрицании закона «исключенного третьего». Этот закон говорит о том, что отрицание ложности некоторого выражения эквивалентно утверждению истинности этого выражения. (Или в принятой символике: $\sim(\sim P) \Leftrightarrow P$, отношение, которое нам уже встречалось ранее.) Наверное, Аристотель был бы очень недоволен, столкнувшись с отрицанием настолько логически «очевидного» факта! С общепринятых позиций здравого смысла закон «исключенного третьего» может рассматриваться как самоочевидная истина: если утверждение о том, что нечто ложно, само неверно, то это нечто должно

[6)] Интуиционизм был назван так потому, что ему предназначалось служить отражением человеческой мысли.

быть непременно справедливым! (На этом законе основана математическая процедура *«доказательства от противного»*, упомянутая на с. 83.) Но интуиционисты считают допустимым отвергать справедливость этого закона. Основная причина здесь в том, что они занимают иную позицию по отношению к понятию *существования*, требуя, чтобы перед признанием существования математического объекта предъявлялось его конкретное (мысленное) построение. То есть, для интуиционалиста «существование» означает «конструктивное существование». В математическом доказательстве, использующем принцип «доказательства от противного», сперва выдвигается некая гипотеза, ложность которой затем устанавливается путем обнаружения противоречий, к которым приводят следствия из этой гипотезы. Эта гипотеза может принимать форму утверждения о том, что математический объект с требуемыми свойствами не существует. Когда это приводит к противоречию, то в *обычной математике* делается вывод о том, что данный объект да, существует. Но подобное доказательство, само по себе, не содержит руководства для *построения* такого объекта. Такое существование для интуициониста существованием отнюдь не является; и именно на этом основании они отказываются признавать закон «исключенного третьего» и процедуру «доказательства от противного». Сам Брауэр был совершенно неудовлетворен таким неконструктивным подходом к понятию существования [7]. Без указания реально осуществимого метода построения, говорил он, такая теория существования будет бессмысленной. В логике Брауэра нельзя сделать заключение о существовании объекта, исходя из ложности утверждения о его несуществовании!

По моему мнению, несмотря на похвальное стремление искать «конструктивное» решение вопроса о математическом существовании, интуиционизм, исповедуемый Брауэром, все же является слишком радикальным. Брауэр впервые опубликовал свои идеи в 1924 году, более чем за десять лет до работ Тьюринга и Черча. Теперь, когда понятие конструктивности — в терминах теории Тьюринга о вычислимости — может изучаться в общепринятых рамках математической философии, уже нет необходимости впадать в крайности, как к тому нас призывает Брауэр. Мы можем исследовать конструктивность как самостоятельный предмет, отдельный от вопроса математического существования. Если мы последуем путем интуиционизма, то будем вынуждены отказаться от использования очень мощных приемов доказательства в математике, заметно ограничивая и лишая силы сам предмет.

Я не хочу излишне подробно останавливаться на разнообразных трудностях и кажущихся абсурдностях, к которым приводит интуиционистский подход; но упоминание некоторых проблем может оказаться полезным. Один из примеров, к которому часто обращается для иллюстрации Брауэр, касается дробной части числа π:

$$3{,}141592653589793\ldots.$$

Существует ли двадцать последовательных семерок где-нибудь в этой части, т. е.:

$$\pi = 3{,}141592653\ldots77777777777777777777\ldots,$$

или же нет? В обычных математических терминах все, что мы можем сказать на сегодняшний день, это то, что они либо существуют, либо нет — и мы не знаем, какая из этих возможностей верна! Казалось бы, вполне безобидное утверждение. Однако правомерность утверждения «последовательность из двадцати семерок либо существует где-то в дробной части числа π, либо нет» будет отвергаться интуиционистами до тех пор, пока не получится установить (некоторым приемлемым с точки зрения интуиционизма конструктивным образом), что такая последовательность действительно существует, или же что такой последовательности нет! Прямого подсчета было бы достаточно для того, чтобы доказать, что данная последовательность действительно существует в дробной части π, но для доказательства невозможности ее существования потребовалась бы математическая теорема. Пока ни один компьютер не в состоянии просчитать дробную часть π с такой точностью, чтобы определить наличие там искомой последовательности. Можно было бы, с вероятностной точки зрения, предположить ее существование, однако, даже

если бы компьютер вычислял каждую секунду, скажем, по 10^{10} цифр, то для нахождения этой последовательности потребовалось бы предположительно от ста до тысячи лет. Мне представляется гораздо более вероятным, что существование такой последовательности будет однажды установлено скорее математически, чем путем прямых вычислений (возможно, как побочный результат более глобального и интересного исследования) — хотя не исключено, что это будет сделано неприемлемым для интуиционистов способом!

Данная проблема не имеет для математики особого значения и приведена лишь как наглядный пример. Брауэр, с позиций радикального интуиционизма, сказал бы, что в настоящее время утверждение «где-то в дробной части числа π существует двадцать последовательных семерок» не является ни справедливым, ни ложным. Если когда-либо в дальнейшем будет установлен конкретный результат — посредством вычислений или путем (интуиционистского) математического доказательства — то тогда утверждение станет «истинным» или «ложным», соответственно. Сходный пример представляет собой и «последняя теорема Ферма». Вновь, согласно крайнему интуиционизму Брауэра, это утверждение не может быть сегодня признано ни ложным, ни истинным, но возможно, что его значение будет определено в будущем. По-моему, такая субъективность и «конъюнктурность» понятия математической истины просто неприемлема. Действительно, вопрос, будет ли — а если будет, то когда — официально признана «доказанность» некоторого математического результата, является весьма субъективным. Математическая истина не должна подчиняться такому «общественно-зависимому» критерию. Помимо этого, опираться на понятие математической истины, зависящее от времени — это, мягко говоря, наиболее неудобный и неудовлетворительный подход для математики, которую предполагается использовать для достоверного описания физического мира. Не все интуиционисты придерживаются таких радикальных взглядов, как Брауэр. И все же точка зрения интуиционистов является, бесспорно, крайне неудобной, даже когда она родственна идеям конструктивизма. Немногие современные математики строго исповедуют чистый интуиционизм, даже если бы единственной причиной этого была бы его ограниченность относительно типов математических рассуждений, которые он позволяет использовать.

Я коротко описал три основных направления в современной математической философии: формализм, платонизм и интуиционизм. Я не скрываю, что практически целиком отдаю предпочтение платонистской точке зрения, согласно которой математическая истина абсолютна и вечна, является внешней по отношению к любой теории и не базируется ни на каком «рукотворном» критерии; а математические объекты обладают свойством собственного вечного существования, не зависящего ни от человеческого общества, ни от конкретного физического объекта. Я попытался привести аргументы в пользу этой точки зрения в этом и предыдущем разделах, а также в конце третьей главы. Я надеюсь, что читатель готов следовать за мной и далее в этих рассуждениях, которые будут очень важны для понимания многих положений в дальнейшем.

Теоремы геделевского типа как следствие результатов, полученных Тьюрингом

В моем изложении теоремы Геделя я опустил многие детали и к тому же оставил в стороне то, что относилось к неразрешимость вопроса о непротиворечивости системы аксиом и было исторически наиболее важной частью его доказательства. Моя задача состояла не в том, чтобы акцентировать внимание на «проблеме доказуемости непротиворечивости аксиом», столь важной для Гильберта и его современников; я стремился показать, что специфическое утверждение Геделя — которое нельзя ни подтвердить, ни опровергнуть исходя из аксиом и правил вывода рассматриваемой формальной системы — оказывается с очевидностью верным, если опираться в наших рассуждениях на интуитивное понимание смысла применяемых процедур.

Я уже упоминал, что Тьюринг разработал свое доказательство неразрешимости проблемы остановки после изучения работ Геделя. Оба доказательства имеют много общего и, естественно, основные положения из результатов Геделя могут быть непосредственно получены путем использования процедуры Тьюринга. Давайте посмотрим, как это происходит, и как при этом можно несколько иным образом взглянуть на то, что осталось за кулисами теоремы Геделя.

Непременное свойство формальной математической системы заключается в существовании вычислимого способа решить, является ли некоторая строка символов доказательством соответствующего математического утверждения или нет. Весь смысл формализации понятия математического доказательства в конечном счете сводится к тому, чтобы не требовалось никакого дополнительного суждения о состоятельности того или иного типа рассуждений. Необходимо обеспечить возможность проверять полностью механическим и заранее определенным способом, что предполагаемое доказательство и в самом деле является таковым — то есть должен существовать алгоритм для проверки доказательств. С другой стороны, мы не требуем существования алгоритмической процедуры *нахождения* доказательств истинности (или ложности) предлагаемых математических утверждений.

Как оказывается, алгоритм отыскания доказательства внутри произвольной формальной системы *присутствует всегда*, если только система допускает какое-нибудь доказательство. Действительно, мы прежде всего должны предполагать, что наша система формулируется на некотором языке символов, который можно выразить в терминах некоторого конечного «алфавита» символов. Как и ранее, давайте упорядочим наши строки символов *лексикографически*, что, как мы помним, означает расставление в алфавитном порядке строк каждой определенной длины, где все строчки единичной длины идут первыми, за ними следуют (также упорядоченные) строки из двух символов, потом — из трех, и так далее (с. 116).

Тогда мы будем иметь корректно построенные и пронумерованные в соответствии с лексикографической схемой доказательства. Располагая нашим списком доказательств, мы одновременно имеем и перечень всех теорем нашей формальной системы, поскольку теорема — это утверждение, которое стоит в последней строчке списка корректно построенных доказательств. Подобный перечень полностью проверяется непосредственными вычислениями: ведь мы можем рассматривать все строки символов системы — независимо от того, имеют они смысл как доказательства или нет — и начать тестировать нашим алгоритмом первую строчку, чтобы понять, является ли она доказательством, и отбросить ее, если нет; затем мы подобным же образом тестируем вторую строчку и исключаем ее, если и она не является доказательством; потом следует третья строчка, четвертая и так далее. Посредством этого мы в конце концов достигнем строки, содержащей доказательство, если таковая имеется в нашем списке.

Таким образом, если бы Гильберту удалось отыскать свою математическую систему — систему аксиом и правил вывода, достаточно мощную, чтобы позволить решать, путем формального доказательства, вопрос о справедливости или ложности любого математического утверждения, корректно сформулированного в рамках системы, — то тогда существовал бы общий алгоритмический метод выяснения истинности любого такого рассуждения. Почему это так? Потому что, если мы при помощи процедуры, описанной выше, находим искомое утверждение как последнюю строчку некоторого доказательства, то это утверждение автоматически считается доказанным. Если же, напротив, мы находим последнюю строчку, содержащую отрицание нашего утверждения, то мы тем самым доказываем его ложность. Если бы схема Гильберта была полной, то либо одна, либо другая возможность обязательно имела бы место (и если бы система была непротиворечивой, то обе возможности никогда бы не могли быть реализованы одновременно). То есть наша механическая процедура всегда бы прерывалась на некотором шаге и мы бы имели универсальный алгоритм для доказательства истинности или ложности всех утверждений системы. Это находилось бы в противоречии с результатами

Тьюринга, изложенными во второй главе, согласно которым не существует общего алгоритма для доказательства математических утверждений. И, как следствие, мы доказали теорему Геделя о том, что ни одна система наподобие задуманной Гильбертом не может быть полной в обсуждаемом нами смысле.

В действительности теорема Геделя носит более частный характер, поскольку от формальной системы того типа, который рассматривал Гедель, требовалась адекватность по отношению к арифметическим утверждениям, а не математическим утверждениям вообще. Можем ли мы устроить так, чтобы все необходимые операции машины Тьюринга выполнялись только при помощи арифметики? Или, иными словами, могут ли все вычислимые функции натуральных чисел (т. е. рекурсивные, или алгоритмические функции — результаты действия машины Тьюринга) быть выражены в терминах обычной арифметики? На самом деле это так, хотя и не совсем. Нам понадобится одна дополнительная операция, которую мы добавим в систему стандартных правил арифметики и логики (включая кванторы $\exists$ и $\forall$). Эта операция просто выбирает «наименьшее натуральное число такое, что $K(x)$ имеет значение „истина“», где $K(\ )$ — заданная арифметически вычислимая *функция исчисления высказываний*, для которой предполагается *существование* такого числа, т. е. что $\exists x[K(x)]$ является истинным. (Если бы такого числа не было, то наша операция повторялась бы «бесконечно»[7], стараясь обнаружить несуществующее x.) В любом случае, предшествующие рассуждения показывают, что, исходя из результатов Тьюринга, программа Гильберта по сведению целых разделов математики к вычислениям в рамках некоторой формальной системы — невыполнима.

Как оказывается, эта процедура не может с очевидностью установить, что мы имеем утверждение Геделя (наподобие $P_k(k)$), которое *верно*, но внутри системы недоказуемо. Однако, если вспомнить доказательство, приведенное в главе 2 и показывающее, «как „перехитрить“ алгоритм» (с. 86), то мы увидим, что можно сделать нечто похожее и в этом случае. В том доказательстве мы смогли выяснить, что для любого алгоритма, определяющего момент остановки машины Тьюринга, можно придумать такое действие машины, которое не прекращается, хотя алгоритм — в отличие от нас — «увидеть» это не способен. (Вспомните, что мы требовали от алгоритма корректно информировать нас о моменте, когда машина Тьюринга действительно остановится, хотя мы допускаем, что он может не оповестить нас, если машина на самом деле *не* прекратит свое действие, продолжая работать вечно.) Таким образом, как и в ситуации с теоремой Геделя, у нас есть утверждение (безостановочное действие машины Тьюринга), истинность которого мы можем установить при помощи интуитивного понимания, хотя определенная алгоритмическая процедура нам такой возможности и не дает.

Рекурсивно нумеруемые множества

Существует способ для описания основных результатов, полученных Геделем и Тьюрингом, в графическом виде, на языке теории множеств. Это позволит нам избежать произвольности описания в терминах конкретного символизма или в рамках формальной системы и выделить наиболее существенное. Мы будем рассматривать только множества натуральных чисел (конечные или бесконечные), такие как $\{4,5,8\}$, $\{0,57,100003\}$, $\{6\}$, $\{0\}$, $\{1,2,3,4,\dots,9999\}$, $\{1,2,3,4,\dots\}$, $\{0,2,4,6,8,\dots\}$ и т. п.; или даже все множество $\mathbb{N} = \{0,1,2,3,4\dots\}$, равно как и пустое множество $\varnothing = \{\}$. Нас будут интересовать только вопросы вычислимости, скажем: «Какие множества натуральных чисел могут быть сгенерированы с помощью алгоритма, а какие — нет?»

Чтобы сформулировать такой вопрос, мы можем считать, что каждое отдельное число n обозначает определенную строчку символов некоторой формальной системы.

[7] Вообще говоря, для нас является существенным, чтобы такие неудачные варианты могли реализоваться, тем самым гарантируя нам потенциальную возможность описывать *любую* алгоритмическую операцию. Вспомните, что для описания машин Тьюринга в общем мы должны допустить существование, в частности, машин, которые никогда не останавливаются.

Это будет n-я строка символов, скажем, Q_n, согласно заданному в системе лексикографическому порядку («синтаксически корректных») утверждений. Тогда каждое натуральное число будет представлять некое утверждение. При этом множество *всех* утверждений формальной системы соответствует всему множеству натуральных чисел; а, допустим, *теоремы* этой системы будут составлять некоторое меньшее множество натуральных чисел, скажем, множество P. Однако детали произвольной системы нумерации утверждений для нас несущественны. Все, что нам потребуется для установления соответствия между натуральными числами и утверждениями — это заданный алгоритм получения каждого утверждения Q_n (записанного должным образом в символических обозначениях) из отвечающего ему натурального числа n; и другой алгоритм для получения n из Q_n. Имея эти алгоритмы в своем распоряжении, мы вольны идентифицировать множество натуральных чисел с множеством утверждений конкретной формальной системы.

Давайте выберем формальную систему достаточно непротиворечивую и широкую для того, чтобы включать в себя все действия всех машин Тьюринга — и, более того, «имеющую смысл» с учетом требования «самоочевидной справедливости» ее аксиом и правил вывода. Далее, пусть ряд утверждений $Q_0, Q_1, Q_2, \ldots$ формальной системы имеет доказательства внутри системы. Эти «доказуемые» утверждения будут иметь номера, которые составляют некоторое множество в $\mathbb{N}$ — по сути, это множество P «теорем», рассмотренных выше. Мы уже видели, что существует *алгоритм* для последовательного построения всех утверждений произвольно заданной формальной системы, имеющих доказательства. (Как отмечено ранее, «n-е доказательство» Π_n получается из n алгоритмически. Все, что нам надо — это посмотреть на последнюю строчку n-го доказательства, чтобы найти «n-е утверждение, доказуемое в рамках системы», т. е. n-ю «теорему».) Следовательно, мы имеем алгоритм последовательной генерации элементов P (при которой возможны и повторения, что для нас не важно).

Множество типа P, которое может быть построено с помощью некоторого алгоритма, называется *рекурсивно нумеруемым*. Заметьте, что множество утверждений, ложность которых может быть установлена в рамках системы — т. е. утверждений, чьи отрицания являются справедливыми — точно так же рекурсивно нумеруемо, поэтому мы можем просто нумеровать доказуемые утверждения по мере продвижения, учитывая и их отрицания. Есть большое число других, тоже рекурсивно нумеруемых, подмножеств $\mathbb{N}$, для определения которых нам вовсе необязательно ссылаться на нашу формальную систему. Простыми примерами рекурсивно нумеруемых множеств могут служить множество четных чисел

$$\{0, 2, 4, 6, 8, \ldots\},$$

множество квадратов

$$\{0, 1, 4, 9, 16, \ldots\}$$

и множество простых чисел

$$\{2, 3, 5, 7, 11, \ldots\}.$$

Очевидно, мы можем построить любое из этих множеств при помощи алгоритма. Для каждого из этих трех примеров будет справедливо следующее свойство: *дополнительное* по отношению к рассматриваемому множество (состоящее из всех натуральных чисел, не входящих в исходное множество) является также рекурсивно нумеруемым. Дополнительными по отношению к вышеприведенным множествам будут, соответственно:

$$\{1, 3, 5, 7, 9, \ldots\}, \quad \{2, 3, 5, 6, 7, 8, 10, \ldots\}$$

и

$$\{0, 1, 4, 6, 8, 9, 10, 12, \ldots\}.$$

Было бы достаточно просто указать алгоритм и для этих дополнительных множеств. Конечно же, мы можем выяснить алгоритмическим путем, является ли произвольное натуральное число n четным, квадратом натурального числа или простым числом, соответственно. Это дает нам алгоритм для построения обоих множеств, поскольку мы можем перебирать все натуральные числа и для каждого из них решать, принадлежит ли оно к определенному множеству или же к его дополнению. Множество, которое обладает

свойством рекурсивной нумеруемости вместе со своим дополнением, называется *рекурсивным*. Очевидно, что дополнительное по отношению к рекурсивному множество также будет рекурсивным.

А существуют ли множества, которые рекурсивно нумеруемы, но рекурсивными, тем не менее, не являются? Давайте на минутку задумаемся над тем, какие следствия могут вытекать из подобного свойства. Поскольку элементы такого множества могут быть получены алгоритмическим путем, мы имели бы способ решить, принадлежит ли некоторый элемент — который, мы предполагаем, да, принадлежит множеству, — рассматриваемому множеству или нет. Все, что от нас требуется, — это запустить алгоритм и прогонять его через все элементы множества до тех пор, пока он не найдет элемент, который мы ищем. Теперь давайте предположим, что искомый элемент *не* принадлежит данному множеству. В таком случае использование нашего алгоритма ничего не даст: он будет работать вечно, будучи не в состоянии прийти к решению. В этом случае нам потребуется алгоритм для построения *дополнительного* по отношению к исходному множества. Если *этот* алгоритм сможет обнаружить искомый элемент, то мы будем точно знать, что он не входит в состав исследуемого множества. Имея на вооружении оба алгоритма, мы так или иначе найдем данный элемент путем поочередного применения этих алгоритмов. Однако, такой благоприятный исход будет иметь место только в случае *рекурсивного* множества. Мы же предполагаем, что мы рассматриваем множество рекурсивно нумеруемое, но при этом не рекурсивное, т. е. наш предполагаемый алгоритм для построения дополнительного множества просто не существует! Таким образом, мы имеем курьезную ситуацию, когда можно определить, включен ли элемент в множество при условии, что он ему наверняка принадлежит; но в то же время нельзя гарантировать, что мы сможем это сделать посредством какого бы то ни было алгоритма для элементов, которые множеству *не* принадлежат.

Может ли возникнуть такая ситуация в реальности? Есть ли и вправду рекурсивно нумеруемые множества, не являющиеся рекурсивными? А как насчет множества P? Имеет ли *это* множество свойство рекурсивности? Мы знаем, что оно рекурсивно нумеруемо, так что нам остается только выяснить, будет ли также дополнительное к нему множество обладать этим свойством. Оказывается, что нет! Как мы можем сделать такой вывод? А давайте вспомним, что наряду с остальными операциями в нашей формальной системе разрешены и действия машин Тьюринга. Если мы обозначим n-ю машину Тьюринга через T_n, то выражение

«$T_n(n)$ останавливается»

— это утверждение — запишем его как $S(n)$, — которое мы можем сформулировать в рамках нашей системы для любого n. Утверждение $S(n)$ будет справедливым для одних значений n, и ложным — для остальных. Множество всех $S(n)$, образованное перебором натуральных чисел $0, 1, 2, 3, \ldots$, будет представлено некоторым подмножеством $\mathbb{N}$ — скажем, S. Теперь учтем фундаментальный результат Тьюринга (глава 2, с. 83), который говорит о том, что не существует алгоритма, способного установить факт «$T_n(n)$ не останавливается» как раз в тех случаях, когда она действительно не останавливается. Это означает, что множество, состоящее из *отрицаний* $S(n)$, не является рекурсивно нумеруемым.

Мы видим, что часть S, принадлежащая P, состоит только из *истинных* $S(n)$. Почему это так? Понятно, что если любое конкретное $S(n)$ доказуемо, то оно должно быть верным (ведь наша формальная система была выбрана так, чтобы иметь «смысл»!), и поэтому часть S, лежащая в P, должна состоять исключительно из *справедливых* утверждений $S(n)$. Более того, ни одно верное утверждение $S(n)$ не должно лежать вне P, ибо, если $T_n(n)$ останавливается, то мы можем отыскать доказательство этому в рамках нашей системы[8].

Теперь предположим, что дополнение P рекурсивно нумеруемо. Тогда у нас был бы алгоритм для построения элементов этого

[8] Доказательство могло бы, в действительности, состоять из последовательности шагов, которые отражали бы действие машины, продолжающееся до ее остановки. Доказательство завершалось бы, как только машина остановится.

дополнительного множества. И мы смогли бы запустить его и пометить каждое утверждение $S(n)$, которое попадает в поле его действия. Это все будут ложные утверждения $S(n)$, так что наша процедура, по сути, обеспечит нам рекурсивную нумерацию множества таких утверждений. Но выше мы установили, что это множество *не нумеруемо* таким образом. Это противоречие показывает, что дополнение P все-таки не может быть рекурсивно пронумеровано; а P, следовательно, *не является рекурсивным*, что и требовалось доказать.

Эти свойства с очевидностью демонстрируют, что наша формальная система не может быть полной: то есть всегда будут существовать утверждения, чью справедливость (или ложность) невозможно доказать в рамках системы. Ведь если предположить, что такие «неразрешимые» утверждения не существуют, то дополнение множества P с необходимостью было бы множеством *опровергаемых* утверждений (все, что недоказуемо, обязано быть опровергаемо). Но мы уже знаем, что опровергаемые утверждения составляют рекурсивно нумеруемое множество, что делает P *рекурсивным*. Однако, P *не* рекурсивно — противоречие, которое доказывает требуемую неполноту. Это основное утверждение теоремы Геделя.

А как насчет подмножества T множества $\mathbb{N}$, которое состоит из *истинных* утверждений нашей формальной системы? Рекурсивно ли T? Или оно только рекурсивно нумеруемо? А его дополнение? Оказывается, что ответ на все эти вопросы — отрицательный. Один из способов установить это — воспользоваться сделанным ранее выводом о невозможности алгоритмически сгенерировать ложные утверждения вида «$T_n(n)$ останавливается». Как следствие, ложные утверждения *в целом* не могут быть получены с помощью алгоритма, поскольку такой алгоритм, в частности, пронумеровал бы все вышеупомянутые ложные «$T_n(n)$ останавливается»-утверждения. Аналогично, и множество всех *истинных* утверждений не может быть построено при помощи алгоритма (так как любой подобный алгоритм легко модифицируется для нахождения ложных утверждений путем отрицания каждого из генерируемых им утверждений). Поскольку, тем самым, истинные утверждения не являются (равно как и ложные) рекурсивно нумеруемыми, то они образуют гораздо более глубокий и сложноорганизованный массив, чем утверждения, имеющие доказательство внутри системы. И это иллюстрирует еще один аспект теоремы Геделя: что понятие математической *истины* только частично досягаемо в рамках любой формальной системы.

Существуют некоторые простые классы истинных арифметических утверждений, которые все же образуют рекурсивно нумеруемые множества. Например, как это нетрудно видеть, истинные утверждения вида

$$\exists\, w, x \ldots, z[f(w, x, \ldots, z) = 0],$$

где $f(\,)$ — некоторая функция, построенная из обычных арифметических операций сложения, вычитания, умножения и возведения в степень, составляют рекурсивно нумеруемые множества [8] (которые я обозначу через A). Пример утверждения такого рода — хотя мы не знаем, верно ли оно — это отрицание последней теоремы Ферма[9], для которой мы можем взять за $f(\,)$ функцию

$$f(w, x, y, z) = \\ = (x+1)^{w+3} + (y+1)^{w+3} - (z+1)^{w+3}.$$

Однако, множество A не является рекурсивным (факт, который *не так* легко установить, хотя он и вытекает из оригинального доказательства Геделя). Значит, мы не имеем никаких алгоритмических средств для выяснения — хотя бы в принципе — истинности или ложности последней теоремы Ферма.

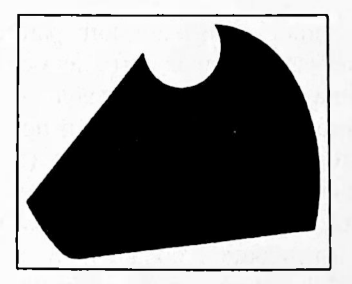


Рис. 4.1. Очень схематичное представление рекурсивного множества

На рис. 4.1 я попытался схематически представить рекурсивное множество как

[9] См. сноску 5 на с. 82. — *Прим. ред.*

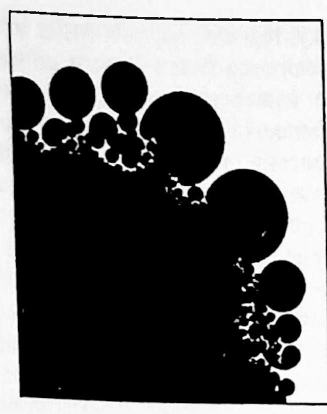


Рис. 4.2. Очень схематичное представление рекурсивно нумеруемого множества (темная область), которое не является рекурсивным. Здесь светлая область определяется только по «остаточному принципу», когда удаляется темная часть, построенная при помощи вычислений; а установить путем прямых вычислений, принадлежит ли заданная точка белой области, нельзя

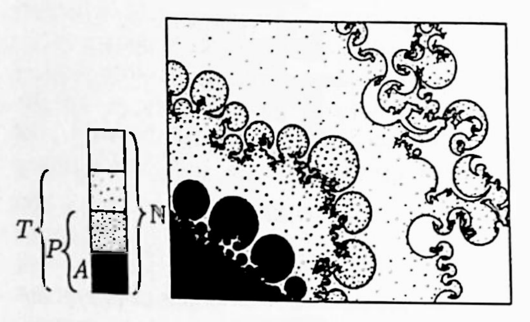


Рис. 4.3. Очень схематичное представление различных множеств утверждений. Множество P утверждений, доказуемых в рамках системы, является, как и A, рекурсивно нумеруемым, но не рекурсивным. Множество T истинных утверждений даже не рекурсивно нумеруемо

фигуру с простой и изящной границей, так что кажется, что определить непосредственно принадлежность произвольной точки этому множеству — дело несложное. Каждая точка на рисунке соответствует некоторому натуральному числу. При этом дополнительное множество также представлено в виде просто выглядящей области на плоскости. На рис. 4.2 я постарался изобразить рекурсивно нумеруемое, но *не рекурсивное* множество в виде области со сложной границей, где подразумевается, что множество с одной стороны границы, — той, что рекурсивно нумеруема — должно выглядеть проще, чем с другой. Фигуры очень схематичны и не претендуют на какую бы то ни было «геометрическую аккуратность». И конечно же, не стоит придавать большого значения тому, что эти рисунки изображены так, как если бы они были расположены на двумерной плоскости!

На рис. 4.3 я схематично обозначил, как расположены области P, T и A внутри множества $\mathbb{N}$.

Является ли множество Мандельброта рекурсивным?

Существенной характеристикой нерекурсивных множеств является их сложноорганизованность. Это свойство должно, в некотором смысле, препятствовать любым попыткам систематизации, которая, в противном случае, привела бы к некоторой «работающей» алгоритмической процедуре. Для нерекурсивного множества не существует общего алгоритмического пути к решению вопроса о принадлежности ему произвольного элемента (или «точки»). В начале третьей главы мы встретились с неким чрезвычайно сложно выглядящим множеством — с множеством Мандельброта. Хотя правила, по которым оно строится, поразительно просты, само множество представляет собой бесконечное разнообразие в высшей степени замысловатых структур. Может ли это быть примером настоящего нерекурсивного множества, явленного глазам смертных?

Читателю, однако, не понадобится много времени, чтобы сообразить, что эта парадигма сложности была создана специально для наших глаз волшебством вычислительных технологий с использованием современных быстродействующих компьютеров. А не являются ли компьютеры истинным воплощением алгоритмических действий? Конечно, это так, но все же мы должны принимать во внимание способ, с помощью которого компьютеры, в действительности, создают эти картинки. Чтобы проверить, принадлежит точка плоскости Аргана — комплексное число c — множеству Мандельброта (закрашено черным) или его дополнению (светлая область), компьютер, начиная

с нуля, применит отображение

$$z \to z^2 + c$$

сначала к $z = 0$, чтобы получить c; потом к $z = c$, чтобы получить $c^2 + c$; затем к $z = c^2 + c$, чтобы получить $c^4 + 2c^3 + c^2 + c$; и так далее. Если эта последовательность $0,\ c,\ c^2 + c,\ c^4 + 2c^3 + c^2 + c,\ \ldots$ остается ограниченной, то соответствующая точка c будет черной; в противном случае — белой. Как машина определяет, что такая последовательность остается ограниченной? В принципе, этот вопрос предполагает наличие информации о том, что происходит после *бесконечного* числа ее элементов! Сама по себе эта задача вычислительными методами не решается. К счастью, существуют способы предсказать исходя уже из конечного числа членов, когда последовательность станет неограниченной. (На самом деле, если последовательность достигает окружности радиуса $1 + \sqrt{2}$ с центром в начале координат, можно с уверенностью сказать, что она будет неограниченной.)

Таким образом, *дополнение* к множеству Мандельброта является, в некотором смысле, рекурсивно нумеруемым. Если комплексное число c расположено в светлой области, то существует алгоритм, подтверждающий этот факт. А как насчет самого множества Мандельброта — темного участка рисунка? Существует ли алгоритм, способный точно установить, что точка, принадлежащая предположительно темному участку, действительно ему принадлежит? Ответ на этот вопрос в настоящее время, похоже, отсутствует [9]. Я справлялся у многих коллег и экспертов, но ни один из них не слышал о подобном алгоритме. Равно как и никто из них не сталкивался с указанием на то, что такого алгоритма не существует. По крайней мере, насколько можно об этом судить, алгоритм для темной области на сегодняшний день неизвестен. Возможно, множество, дополнительное по отношению к множеству Мандельброта, действительно является примером рекурсивно нумеруемого, но не рекурсивного множества!

Прежде чем исследовать дальше это предположение, необходимо будет обсудить некоторые моменты, которые я ранее опускал. Эти вопросы будут довольно важны для нас в дальнейших рассуждениях по поводу вычислимости в физике. Я хотел бы заметить, что, на самом деле, я был несколько неточен в предшествующем изложении. Я применял такие понятия, как «рекурсивно нумеруемый» и «рекурсивный», к множествам точек в плоскости Аргана, т. е. множествам комплексных чисел. Но эти термины могут применяться только лишь для натуральных чисел и других *счетных* множеств. Мы видели в третьей главе (с. 100), что действительные числа не могут быть счетным множеством, равно как, следовательно, и комплексные — ведь любое действительное число может быть рассмотрено как частный случай некоторого комплексного числа с нулевой мнимой частью (с. 103). В действительности существует такое же «количество» комплексных чисел, как и действительных, а именно «C». (Чтобы установить взаимно-однозначное соответствие между комплексными и действительными числами, можно, грубо говоря, просто взять действительную и мнимую части комплексного числа (записанные в десятичной форме) и перемешать через одну поразрядно цифры из мнимой части с цифрами из вещественной, образуя, тем самым, действительное число: тогда, например, $3{,}6781\ldots + i512{,}975\ldots$ будет соответствовать действительному числу $50\,132{,}6977851\ldots$.)

Дабы избежать этой проблемы, можно было бы ограничиться только *вычислимыми* комплексными числами, так как мы еще в третьей главе видели, что вычислимые действительные числа — а значит, и соответствующие им комплексные — являются счетными. Однако здесь кроется одна принципиальная трудность: не существует алгоритма, с помощью которого можно было бы сравнивать два вычислимых числа, полученных алгоритмически! (Мы можем алгоритмическим образом составить их разность, но мы не в состоянии будем выяснить, равна она нулю или нет. Представьте себе два алгоритма, которые генерируют цифры $0{,}99999\ldots$ и $1{,}00000\ldots$, соответственно; мы никогда не узнаем, продолжаются ли нули и девятки в них до бесконечности — так, что числа оказываются равными — или же где-то в дробной части того или другого числа могут появиться иные цифры, делая эти

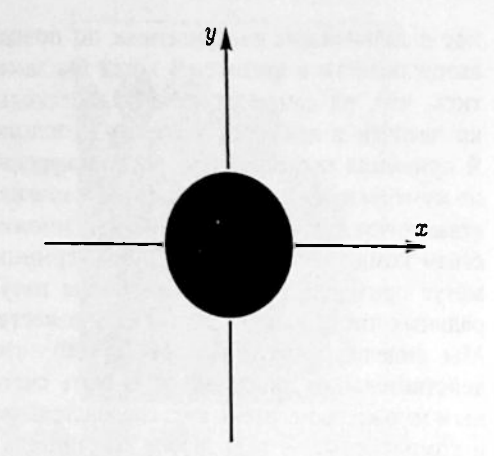


Рис. 4.4. Единичный круг, безусловно, должен рассматриваться как рекурсивное множество, но это требует определенных соглашений

числа неравными.) Таким образом, мы, возможно, никогда не сможем определить, равны ли между собой такие числа. Как следствие этого — наша неспособность решить даже в таком простом случае как единичный круг в плоскости Аргана (множество точек, лежащих на расстоянии не большем единицы от начала координат — черная фигура на рис. 4.4), лежит ли комплексное число в этом круге или нет. Трудность возникает не с точками, лежащими внутри или снаружи, а именно с точками на самой границе круга — то есть на самой единичной окружности. Эта окружность рассматривается по условию как часть круга. Предположим, что нам уже предоставлен в распоряжение алгоритм для получения цифр вещественной и мнимой частей некоторого комплексного числа. Если мы предполагаем, что это комплексное число лежит на единичной окружности, то мы не можем с необходимостью подтвердить этот факт. Не существует алгоритма, чтобы установить, является ли вычислимое число

$$x^2 + y^2$$

равным единице, что служит критерием для принадлежности комплексного числа $x + iy$ данной единичной окружности.

Очевидно, это совсем не то, что нам нужно. Единичный круг, безусловно, *должен* рассматриваться как рекурсивное множество. Едва ли найдется сколь-нибудь значительное число множеств, более простых, чем единичный круг! Чтобы обойти эту проблему, одним из способов может быть *игнорирование* границы. Ведь для точек, лежащих внутри (или снаружи), безусловно существует алгоритм, устанавливающий этот факт. (Можно просто последовательно генерировать цифры числа x^2+y^2, и, в конце концов, мы найдем цифру, отличную от 9 в дробной части 0,9999... или отличную от 0 — в дробной части 1,00000... .) В этом смысле единичный круг *является* рекурсивным. Но этот подход чрезвычайно неудобен для математики, поскольку там часто возникает необходимость ссылаться в рассуждениях на то, что происходит *именно* на границах. С другой стороны, вполне возможно, что такая точка зрения окажется применимой в области физики. Позднее нам еще придется вернуться к этому вопросу.

Существует другой метод, имеющий непосредственное отношение к данному вопросу, который не предполагает вообще обращения к вычислимым комплексным числам. Вместо того, чтобы пытаться пронумеровать комплексные числа внутри или снаружи рассматриваемого множества, мы просто будем вызывать алгоритм, который для любого *наперед заданного* комплексного числа будет определять, принадлежит оно нашему множеству или же его дополнению. Говоря «наперед заданный», я подразумеваю, что для каждого числа, которое мы рассматриваем, нам некоторым — быть может, «волшебным» — образом известны цифры мнимой и вещественной части, одна за другой, и в таком количестве, сколько нам нужно. Я не требую, чтобы существовал алгоритм, известный или неизвестный, для *нахождения* этих цифр. Множество комплексных чисел считалось бы «рекурсивно нумеруемым», если бы существовал хотя бы единственный алгоритм такой, что для любой заданной ему вышеуказанным образом последовательности цифр он бы говорил «да» после конечного числа шагов тогда и только тогда, когда комплексное число действительно принадлежит этому множеству. Оказывается, что как и в случае подхода, предложенного выше, эта точка зрение также «игнорирует» границы. Следовательно, внутренняя и внешняя области единичного диска будут каждая по отдельности счи-

таться рекурсивно нумеруемыми в указанном смысле, тогда как сама граница — нет.

Для меня совершенно не очевидно, что какой-либо из этих методов дает то, что нам нужно [10]. Философия «игнорирования границ», будучи приложенной к множеству Мандельброта, может привести к потере большого числа тонких моментов. Одна часть этого множества состоит из «клякс» — внутренних областей, а другая — из «усиков». Наибольшие сложности при этом связаны, видимо, с «усиками», которые могут «извиваться» самым причудливым образом. Однако, «усики» не принадлежат внутренней части множества, и, тем самым, они были бы проигнорированы, используй мы любой из двух вышеприведенных подходов. Но даже при таком допущении остается неясность, можно ли считать множество Мандельброта рекурсивным в том случае, когда рассматриваются только «кляксы». Похоже, что вопрос этот связан с некоторым недоказанным предположением, касающимся самого множества, а именно: является ли оно, что называется, «локально связным»? Я не собираюсь здесь разбирать значение этого понятия или его важность для данного вопроса. Я хочу просто показать, что существует ряд трудностей, которые вызывают неразрешенные на сегодняшний день вопросы, касающиеся множества Мандельброта, чье решение — первоочередная задача для некоторых современных математических исследований.

Существуют также и другие подходы, которые могут использоваться с тем, чтобы обойти проблему несчетности комплексных чисел. Вместо того, чтобы рассматривать все вычислимые комплексные числа, можно ограничиться только подмножеством таких чисел, для любой пары которых можно вычислительным путем установить их равенство. Простым примером такого подмножества могут служить *«рациональные»* комплексные числа, у которых как мнимая, так и вещественная части могут быть представлены рациональными числами. Я не думаю, однако, что это дало бы многое в случае «усиков» множества Мандельброта, поскольку такая точка зрения накладывает очень значительные ограничения. Более удовлетворительным могло бы оказаться рассмотрение *алгебраических* чисел — тех комплексных чисел, которые являются алгебраическими решениями уравнений с целыми коэффициентами. Например, все решения z уравнения

$$129z^7 - 33z^5 + 725z^4 + 16z^3 - 2z - 3 = 0$$

— это алгебраические числа. Такие числа будут счетными и вычислимыми, и задача проверки двух из них на равенство будет решаться путем прямого вычисления. (Как выясняется, многие из них будут лежать на границе единичного круга и «усиков» множества Мандельброта.) И мы можем по желанию рассматривать вопрос о рекурсивности множества Мандельброта в терминах этих чисел.

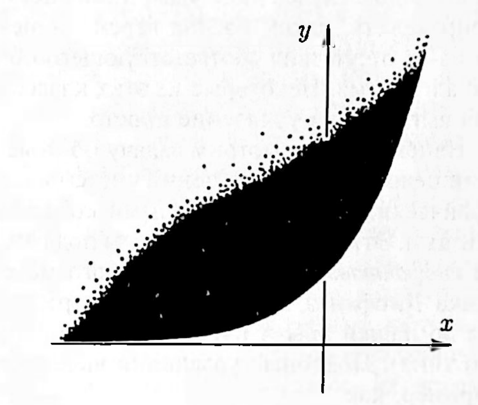


Рис. 4.5. Множество, определенное экспоненциальным соотношением $y \geqslant e^x$, должно также рассматриваться как рекурсивное

Возможно, что алгебраические числа оказались бы подходящим инструментом для двух обсуждаемых нами множеств, но они не снимают все наши трудности в общем случае. Пусть мы рассматриваем множество (темная область на рис. 4.5), определяемое неравенством

$$y \geqslant e^x,$$

где $x + iy (= z)$ — точка в плоскости Аргана. Внутренняя часть множества, равно как и внутренняя часть его дополнения, будут рекурсивно нумеруемыми в соответствии с любой из вышеизложенных точек зрения, но (как следует из знаменитой теоремы Ф. Линдеманна, доказанной в 1882 году) граница, $y = e^x$, содержит только одну алгебраическую точку, а именно точку $z = i$. В этом случае алгебраические числа никак не могут нам помочь при исследовании алгоритмической по своей природе границы!

Несложно определить другой подкласс вычислимых чисел, которые будут подходить в данном конкретном случае, но при этом все равно останется ощущение, что правильный подход нами до сих пор так и не был найден.

Некоторые примеры нерекурсивной математики

Существует немало областей математики, где возникают проблемы нерекурсивного характера. Это означает, что мы можем сталкиваться с задачами, ответ к которым в каждом случае либо «да», либо «нет», но определить, какой из них верен, — нельзя из-за отсутствия соответствующего общего алгоритма. Некоторые из этих классов задач выглядят на удивление просто.

Например, рассмотрим задачу об отыскании целочисленных решений системы алгебраических уравнений с целыми коэффициентами. Эти уравнения известны под именем *диофантовых* (в честь греческого математика Диофанта, который жил в третьем веке до нашей эры и изучал уравнения такого типа). Подобные уравнения выглядят, например, как

$$z^3 - y - 1 = 0,$$
$$yz^2 - 2x - 2 = 0,$$
$$y^2 - 2xz + z + 1 = 0,$$

и задача состоит в том, чтобы определить, могут ли они быть решены в целых x, y, z. Оказывается, что в этом конкретном случае существует тройка целых чисел, дающая решение этой системы:

$$x = 13, \quad y = 7, \quad z = 2.$$

Но для произвольной системы диофантовых уравнений никакого алгоритма не существует[10]. Арифметика Диофанта, несмотря на простоту входящих в нее выражений, является частью неалгоритмической математики!

(Несколько менее тривиальным является пример *топологической эквивалентности* многообразий. Я упоминаю об этом только вкратце, ибо в главе 8 будут рассматриваться вопросы, имеющие к данному определенное отношение. Чтобы понять, что такое «многообразие», представьте для начала петлю, которая является многообразием в одном измерении; затем представьте замкнутую поверхность — многообразие в двух измерениях. Далее попробуйте представить некую «поверхность», имеющую *три* и более измерений. «Топологическая эквивалентность» двух многообразий означает, что одно из них может быть деформировано в другое путем непрерывных преобразований — без разрывов и склеек. Так, сфера и поверхность куба являются топологически эквивалентными, хотя они не эквивалентны поверхности кольца или чашки с ручкой — хотя последние топологически эквивалентны друг другу. При этом для *двумерных* многообразий существует алгоритм, позволяющий определить, эквивалентны ли произвольные два многообразия друг другу или нет — в сущности, заключающийся в подсчете «ручек», которые имеет каждая из поверхностей. Для случая трех измерений вопрос о существовании такого алгоритма на момент написания книги остается без ответа; однако для четырех и более измерений уже известно, что такого алгоритма быть *не может*. Возможно, четырехмерный случай имеет некое отношение к физике, поскольку согласно теории общей относительности Эйнштейна пространство и время совместно образуют четырехмерное многообразие (см. главу 5, с. 183). Герох и Хартли в 1986 году высказали предположение о том, что свойство неалгоритмичности может иметь отношение к «квантовой гравитации» (см. также главу 8).)

Давайте теперь рассмотрим иной тип задач, называемых *задачами со словами*[11]. Допустим, у нас есть некий алфавит символов, и мы рассматриваем различные строки этих символов, трактуя их как *слова*. Слова могут сами по себе не иметь никакого значения, но мы должны иметь некоторый (конечный) список «равенств» между ними, которые мы сможем использовать для дальнейшего построения таких «равенств». Это делается путем подстановки слов из исходного списка в другие (как правило, более длинные) слова, которые содержат их в виде

---

[10] Это дает (отрицательный) ответ на десятую проблему Гильберта, упомянутую на с. 64 (см., например, Дэвлин [1988]). Здесь количество переменных неограниченно. Однако, известно, что для выполнения свойства неалгоритмичности достаточно и девяти.

составных частей. Каждая такая часть может быть заменена на равную ей в соответствии с используемым списком. Тогда для любой данной пары слов мы должны решить задачу об их равенстве согласно этим правилам.

В качестве примера мы можем взять для нашего исходного списка, скажем, такие равенства:

EAT = AT

ATE = A

LATER = LOW

PAN = PILLOW

CARP = ME.

Отсюда мы можем, например, вывести

LAP = LEAP,

используя последовательные замены из второго, первого и снова второго соотношения из нашего исходного листа:

LAP = LATEP = LEATEP = LEAP.

Проблема теперь заключается в том, чтобы выяснить, сможем ли мы для любой наперед заданной пары слов осуществить вышеописанным образом переход от одного из них к другому? Можем мы перейти от CATERPILLAR к MAN, или, скажем, от CARPET — к MEAT? Ответ в первом случае оказывается утвердительным, тогда как во втором — отрицательным. Когда ответ утвердителен, стандартный путь показать его справедливость заключается в построении цепочки равенств, где каждое из слов получается из предыдущего с учетом допустимых соотношений. Итак, имеем (обозначая буквы, назначенные к замене, жирным шрифтом, а только что измененные — курсивом):

CATERPILLAR = C*A*RPILLAR =

= *CARPIL**LATE**R* = CARPIL*LOW* =

= CAR***PAN*** = *MEAN* = *M**EATE**N* =

= M***AT***EN = M*A*N.

Как мы можем утверждать, что посредством разрешенных подстановок невозможно получить MEAT из CARPET? Для демонстрации этого факта придется подумать чуть больше, однако показать это не так уж сложно, причем множеством разных способов. Простейшим представляется следующий: в каждом «равенстве» из нашего списка число букв A плюс число букв W плюс число букв M с каждой стороны одинаково. Значит, общая сумма указанных букв не может меняться в процессе преобразования по допустимым нашим списком правилам. Однако, для CARPET эта сумма равна 1, а для MEAT — 2. Следовательно, не существует способа получить из первого слова второе при помощи вышеприведенного списка равенств.

Заметьте, что когда два слова «равны», мы можем показать это, просто приведя допустимую формальную строчку символов, построенную с помощью заданных нами правил; тогда как в случае их «неравенства» мы должны прибегать к рассуждениям *об* этих самых правилах. Существует четкий алгоритм, который мы можем использовать для установления «равенства» между двумя словами в том случае, когда они действительно «равны». Все, что нам требуется, это составить лексикографический перечень всех возможных последовательностей слов, и потом вычеркнуть из этого списка любую строчку, где имеется пара слов, в которой последующее нельзя получить из предыдущего при помощи какого бы то ни было правила из исходного списка. Оставшиеся последовательности дадут нам набор всех искомых «равенств» между словами. Однако, в общем случае нет такого явного алгоритма для случая, когда два слова «неравны», и нам, возможно, пришлось бы применить «интеллект» для установления этого факта. (Конечно же, мне потребовалось некоторое время, прежде чем я заметил описанный выше «трюк», при помощи которого доказал, что CARPET и MEAT «неравны». А для другого примера «трюк» мог бы понадобиться совершенно иной. Кстати, интеллект помогает — хотя и не обязательно — и в случае, когда необходимо установить *существование* некоторого «равенства».)

В действительности, для нашего *конкретного* списка из пяти «равенств», которые составляют исходный список в рассмотренном выше примере, привести алгоритм, устанавливающий «неравенство» в случае, когда два слова и вправду «неравны» — не так уж сложно. Однако, чтобы *отыскать* алгоритм в этом случае, потребовалась

изрядная работа интеллекта! И, конечно, оказывается, что не существует единого универсального алгоритма для *всех* возможных вариантов исходного списка. Общая задача со словами принадлежит области нерекурсивной математики!

Существуют даже определенные варианты выбора исходного списка, для которых нет алгоритма решения задачи сравнения двух слов. Один из них дается таким набором:

AH = HA

OH = HO

AT = TA

OT = TO

TAI = IT

HOI = IH

THAT = ITHT.

(Этот список взят из списка, предложенного Григорием Цейтиным и Даной Скотт в 1955 году (см. Гарднер [1958]).) Таким образом, эта частная задача со словами служит примером нерекурсивной математики в том смысле, что, используя такой исходный список, мы не можем алгоритмическим путем решить, «равны» два наперед заданных слова или нет.

Общая задача со словами возникает как следствие рассмотрения формализованной математической логики («формальных систем» и т. п., в соответствии с обсуждаемым ранее). Исходный список выполняет роль системы аксиом, а правила замены слов — правил вывода. Доказательство нерекурсивности задачи со словами вытекает из подобных рассуждений.

В качестве последнего примера задачи из области нерекурсивной математики давайте рассмотрим вопрос о покрытии Евклидовой плоскости многоугольниками, разнообразие форм которых ограничено, а сам вопрос при этом ставится так: можем ли мы выложить всю плоскость полностью, без разрывов и нахлестов, используя фигуры только данных нам форм? Такая укладка фигур называется *замощением* плоскости. Мы знаем, что такое замощение возможно при помощи только квадратов, только равнобедренных треугольников или только правильными шестиугольниками (как изображено на рис. 10.2 на с. 371), но невозможно,

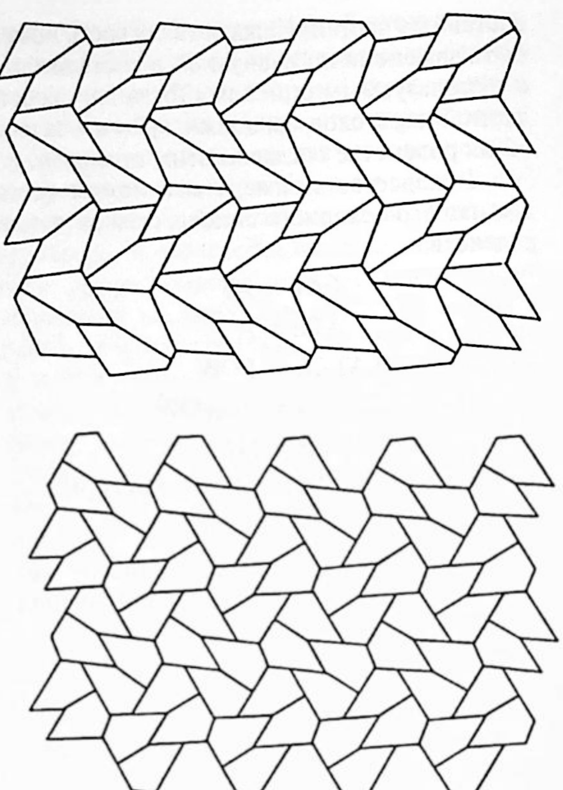


Рис. 4.6. Два примера периодического замощения плоскости фигурой одной формы (предложены Марджори Райс (Marjorie Rice) в 1976 году)

если использовать только правильные пятиугольники. Многими иными фигурами, такими, как два *неправильных* пятиугольника на рис. 4.6, также можно выложить плоскость. Замощение фигурами *двух* форм может стать более хитроумной задачей. Два простых примера даны на рис. 4.7. Все эти замощения являются периодическими; это означает, что они в точности повторяются по всей плоскости в двух независимых направлениях. На языке математики мы бы сказали, что существует *параллелограмм периодов* — параллелограмм, который, будучи неким образом выделен и затем повторен снова и снова в двух направлениях, параллельных его сторонам, даст в результате заданный узор покрытия. На рис. 4.8 представлен пример, где периодическое покрытие слева состоит из «плиток» в форме шипов, а справа указан соответствующий параллелограмм периодов.

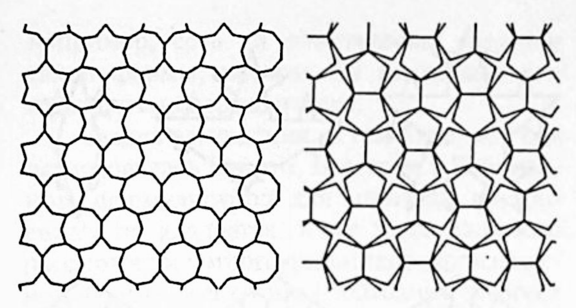


Рис. 4.7. Два примера периодического замощения плоскости фигурами двух форм

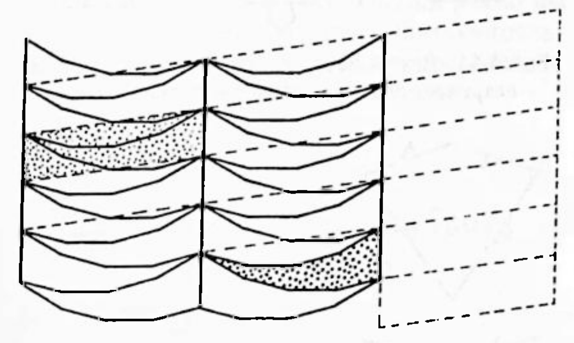


Рис. 4.8. Периодическое замощение и его параллелограмм периодов

С другой стороны, существует множество типов замощений плоскости, которые *не являются* периодическими.

Рис. 4.9 изображает три непериодических «спиральных» замощения из таких же шиповидных «плиток», как и на рис. 4.8. Эта форма «плиток», известная как «универсальная» (по вполне понятным причинам!), была предложена Б. Грюнбаумом и Дж. К. Шепардом [1981, 1987] на основании форм, изученных Х. Фодербергом. Обратите внимание, что универсальная форма позволяет замостить плоскость *как* периодически, *так и* непериодически. Это свойство характерно и для многих других форм единичных «плиток» и наборов «плиток». А могут ли существовать «плитки» (или конечные наборы «плиток»), которые бы покрывали плоскость *только* непериодически? Ответ на этот вопрос будет «да». На рис. 4.10 я изобразил сконструированный американским математиком Рафаэлем Робинсоном набор из фигур шести различных форм, которым можно замостить всю плоскость, но только непериодическим образом.

Небесполезно было бы сделать историческое отступление и посмотреть, как

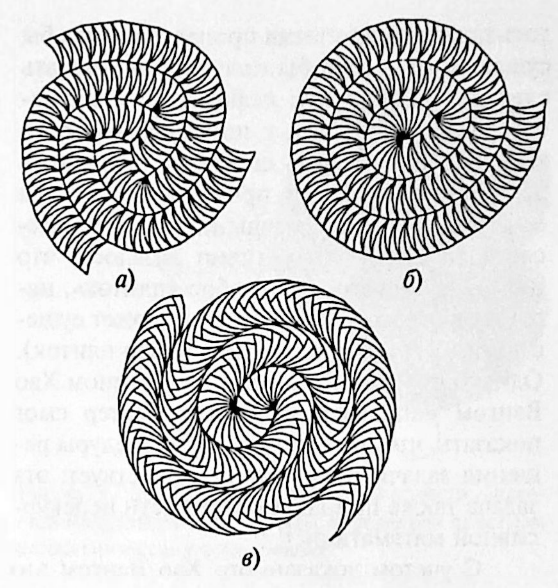


Рис. 4.9. Три непериодических «спиральных» замощения из таких же «универсальных» плиток, как и на рис. 4.8

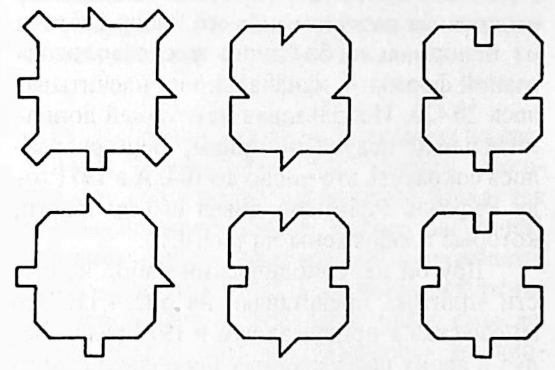


Рис. 4.10. Набор Рафаэля Робинсона из шести плиток, который покрывает плоскость только непериодически

появился это непериодический набор (см. Грюнбаум, Шепард [1987]). В 1961 году американский логик китайского происхождения Хао Ванг поставил вопрос о существовании процедуры для решения задачи замощения, или, иными словами, о нахождении алгоритма, который позволил бы выяснить возможность замощения всей плоскости с помощью конечного набора многоугольников различной формы![11] Ему уда-

[11] В действительности Хао Ванг занимался несколько иной проблемой — с квадратными «плитками», не вращаемыми и с совпадающими по цвету сторонами, — но эти особенности для нас здесь не важны.

лось показать, что такая процедура могла бы существовать, если бы получилось доказать следующую гипотезу: любой конечный набор разных «плиток», с помощью которого можно каким-нибудь способом выполнить замощение плоскости, пригоден также и для ее периодического замощения. Мне думается, в то время интуитивно казалось, что не может существовать набор «плиток», нарушающий это условие (т. е. не может существовать «непериодический» набор плиток). Однако в 1966 году, следуя в указанном Хао Вангом направлении, Роберт Бергер смог показать, что, на самом деле, процедуры решения задачи покрытия *не* существует: эта задача также принадлежит области нерекурсивной математики![12]

С учетом доказанного Хао Вангом это означало, что хотя бы один непериодический набор «плиток» должен существовать; и Бергер смог построить первый такой набор. Однако, из-за сложности выбранного им способа рассуждений, его набор состоял из ненормально большого числа «плиток» разной формы — изначально их насчитывалось 20 426. Использовав некоторый дополнительный искусный прием, Бергеру удалось сократить это число до 104. А в 1971 году Рафаэль Робинсон довел его до шести, которые изображены на рис. 4.10.

Другой непериодический набор из шести «плиток» представлен на рис. 4.11. Это множество я придумал сам в 1973 году, следуя в своих рассуждениях несколько отличным путем. (Я вернусь к этой теме в главе 10, где на рис. 10.3, с. 371, изображен массив, покрытый такими «плитками».) После того, как я познакомился с «шестиплиточным» набором Робинсона, я начал думать о том, как сократить их число; и путем различных манипуляций с разрезаниями и склеиванием я, в конечном счете, смог довести количество «плиток» до двух. Две альтернативные схемы представлены на рис. 4.12. Узоры, которые получаются в результате полного замощения и имеющие с необходимостью непериодическую структуру, обладают рядом замечательных свойств, в том числе — кажущейся невозможной с точки зрения кристаллографии квазипериодической симметрией с осью пятого порядка. К этому вопросу я вернусь позднее.

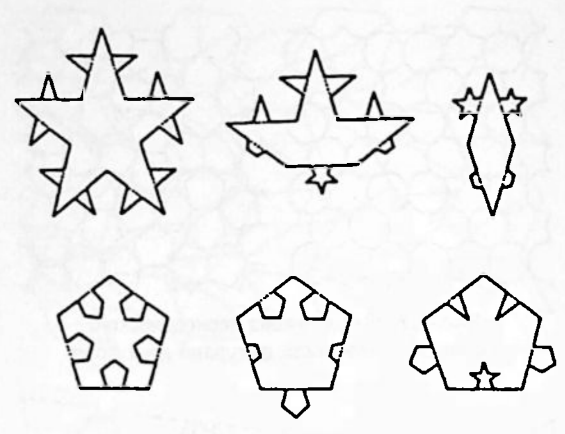


Рис. 4.11. Другой набор из шести плиток, который покрывает плоскость только непериодически

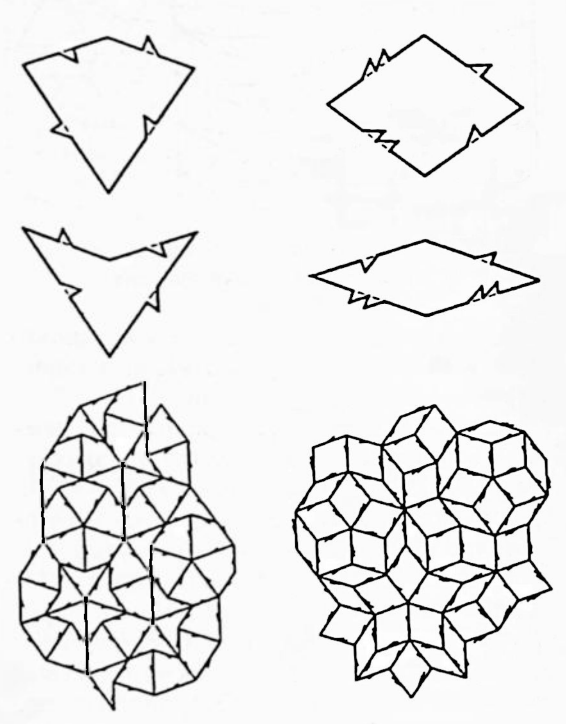


Рис. 4.12. Две пары плиток, которые покрывают плоскость только непериодически («плитки Пенроуза»). Также показано замощение плоскости каждой из этих пар

Вероятно, это покажется удивительным, что такая очевидно «тривиальная» область математики, как замощение плоскости конгруэнтными «плитками», которая выглядит не более серьезно, чем «детская игра», на самом деле является частью нерекурсивной математики. В действительности эта область содержит множество трудных и не решенных пока задач. Пока неизвестно,

например, есть ли *единственная «плитка»* такой формы, которая бы покрывала всю плоскость непериодически.

Задача замощения, в том виде, как она исследовалась Вангом, Бергером и Робинсоном, формулируется для «плиток», построенных на квадратах. Я же здесь допускаю рассмотрение многоугольников произвольной формы, и поэтому необходимо наличие какого-нибудь способа изображения каждой из «плиток», поддающегося адекватному вычислению. Одним из таких путей могло бы быть представление вершин «плиток» точками плоскости Аргана, которые превосходно задаются алгебраическими числами.

Похоже ли множество Мандельброта на нерекурсивную математику?

Давайте теперь вернемся к нашей предшествующей дискуссии о множестве Мандельброта. Я буду для наглядности предполагать, что это множество является в некотором смысле нерекурсивным. Поскольку его дополнение рекурсивно нумеруемо, то, как следствие, само оно таковым быть не может. Я думаю, что форма множества Мандельброта может кое-чему научить нас о том, что касается природы нерекурсивных множеств и нерекурсивной математики.

Посмотрим еще раз на рис. 3.2, с которым мы встретились в третьей главе. Заметьте, что большая часть множества вписывается в сердцевидную фигуру, которую я обозначил на рис. 4.13 через A. Эта фигура называется *кардиоида* и ее внутренняя область может быть определена математически как множество точек c плоскости Аргана, которые удовлетворяют равенству

$$c = z - z^2,$$

где z — комплексное число, чье расстояние до центра координат меньше 1/2. Это множество является, с очевидностью, рекурсивно нумеруемым в смысле существования алгоритма, который для произвольной точки внутренней области фигуры умеет подтверждать ее принадлежность этой самой

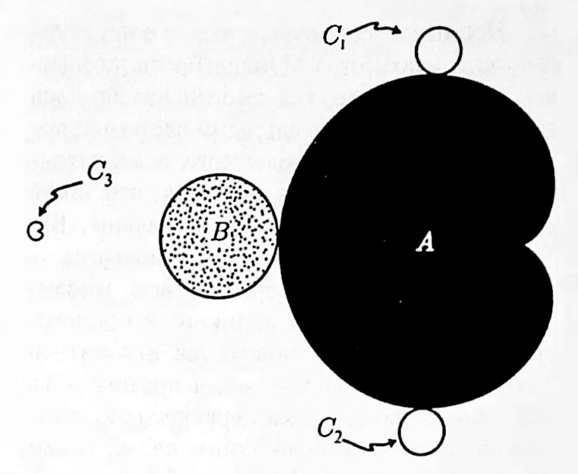


Рис. 4.13. Бо́льшая часть внутренней области множества Мандельброта может быть определена простыми алгоритмическими уравнениями

области. Этот алгоритм легко получается из указанной выше формулы.

Теперь рассмотрим дисковидную фигуру слева от основной кардиоиды (область B на рис. 4.13). Ее внутренняя часть представляет собой множество точек

$$c = z - 1,$$

где z — удалено от начала координат на расстояние меньше 1/4. Эта область, несомненно, является внутренностью диска, так как представляет собой множество точек, лежащих внутри правильной окружности. И, опять же, эта область является рекурсивно нумеруемой в принятом нами смысле. А как насчет других «бородавок» на кардиоиде? Возьмем две следующие по величине «бородавки». Это практически круглые «кляксы», располагающиеся примерно наверху и внизу кардиоиды на рис. 3.2 и которые на рис. 4.13 обозначены через C_1 и C_2. Они могут быть описаны как множество

$$c^3 + 2c^2 + (1 - z)c + (1 - z)^2 = 0,$$

где z изменяется в пределах круга радиуса 1/8 с центром в начале координат. Фактически, это уравнение дает нам не только обе эти «кляксы», но и «дочернюю» фигуру кардиоидной формы (основную часть рис. 3.1), которая находится слева на рис. 3.2 и которая обозначена как C_3 на рис. 4.13. И, аналогично, эти области (как порознь, так и вместе) составляют рекурсивно нумеруемые множества благодаря существованию вышеприведенной формулы.

Несмотря на предположение о нерекурсивности множества Мандельброта, сделанное мной вначале, мы смогли разобраться с его наиболее значительными частями с помощью вполне определенного и достаточно простого алгоритма. Кажется, что такой процесс можно продолжать и дальше. Все наиболее очевидные области множества — и, конечно же, подавляющая часть множества (если не все оно целиком) в процентном выражении — поддаются алгоритмическому анализу. Если, как я предполагаю, все множество все-таки нерекурсивно, то те области, которые недоступны для действия алгоритма, должны быть с необходимостью очень «тонкими» и почти «невидимыми». Более того: когда мы найдем такую область, то вероятнее всего мы смогли бы понять, как нам изменить наш алгоритм, чтобы эта область также оказалась в зоне его действия. Однако, после этого найдутся другие области (если мое предположение о нерекурсивности справедливо), еще более труднодоступные из-за тонкости и сложности своей структуры, перед которыми будет бессилен даже наш усовершенствованный алгоритм. И вновь «волшебство» интуиции, искусства и техники, наверное, позволит нам вычленить эту область; но другие в очередной раз ускользнут от нас; и так будет повторяться снова и снова.

Я полагаю, что этот путь не слишком отличается от того, который часто используется в математике для решения трудных и, предположительно, нерекурсивных задач. Многие задачи, с которыми сталкиваются в некоторых специфических областях, часто решаются с помощью простых алгоритмических процедур — процедур, известных, быть может, на протяжении веков. Но некоторые из этих задач могут не поддаться таким методам, и тогда приходится искать более сложные пути к их решению. Такие задачи будут, конечно, сильнее всего интриговать математиков и подталкивать их к развитию все более мощных методов, в основу которых будет закладываться все более и более глубокое интуитивное понимание природы используемых математических объектов. Возможно, в этом есть что-то от того, как мы познаем окружающий нас физический мир.

В задачах покрытия и задачах со словами, рассмотренных выше, можно уже уловить, как применяется подобный подход (хотя это не те области математики, где аппарат развит в достаточной степени). Мы смогли привести очень простое доказательство для того, чтобы показать невозможность трансформации одного слова в другое при помощи установленных правил. Нетрудно вообразить, что более «продвинутые» методы доказательства способны помочь в более сложных случаях. Не исключена вероятность, что эти новые подходы могут быть превращены в алгоритмические процедуры. Мы знаем, что ни одна процедура не может удовлетворять всем примерам задачи со словами, но те из них, которые ускользают из «алгоритмических сетей», должны быть очень тонко и аккуратно сконструированы. Конечно, как только мы *узнаем* принцип построения таких примеров — как только мы будем уверены, что в неком конкретном случае произошла «осечка» алгоритма, — мы сможем усовершенствовать наш алгоритм так, чтобы он включал и этот частный пример. Ускользать могут только пары «неравных» слов, так что, как только мы находим такую «ускользающую» пару, мы можем быть уверены в их «неравенстве» и присовокупить этот критерий к нашему алгоритму. Так наше более глубокое понимание ведет ко все более совершенным алгоритмам!

Теория сложности

Рассуждения о природе, возможности построения, существования и ограничениях алгоритмов, которые я привел в предыдущих главах, были по большей части «нестрогими». Я совсем не касался вопроса о возможности практического применения упоминавшихся алгоритмов. Даже в тех задачах, где существование алгоритмов и возможные способы их построения очевидны, все же может потребоваться довольно много труда для их воплощения в нечто полезное с точки зрения практического использования. Иной раз небольшая догадка или искусный ход могут в значительной степени упростить алгоритм или же многократно увеличить его быстродействие. Техническая сторона этих

вопросов часто бывает очень сложна, и в последние годы в различных направлениях прилагалось много усилий в области построения, понимания и совершенствования алгоритмов — быстро растущем и развивающемся поле деятельности для пытливых умов. Мне представляется не слишком уместным углубляться здесь в тонкости подобных вопросов. Однако, существует довольно много *абсолютных* ограничений общего характера (известных или предполагаемых) на возможное повышение быстродействия алгоритма. Оказывается, что среди алгоритмических по своей природе задач существуют определенные классы проблем, решать которые с помощью алгоритмов несоизмеримо труднее, чем остальные. Такие задачи можно решать только с помощью очень медленных алгоритмов (или, допустим, алгоритмов, требующих чрезмерно больших ресурсов для хранения информации, и т. п.). Теория, в которой рассматриваются подобные вопросы, носит название *теории сложности*.

Теория сложности занимается не столько изучением трудностей, связанных с решением *отдельных* задач, сколько с бесконечными семействами задач, в каждом из которых любая задача может быть решена с помощью одного и того же алгоритма. Различные задачи такого семейства будут отличаться по «размеру», который выражается некоторым натуральным числом n. (Чуть позднее я объясню более подробно, как фактически этот номер n характеризует размер задачи.) Время, требуемое для решения конкретной задачи из рассматриваемого класса, — а вернее, количество элементарных шагов, — дается некоторым числом N, зависящим от n. Для определенности договоримся, что N — это наибольшее число шагов среди всех задач данного размера n, которое может понадобиться алгоритму для решения. При этом, с ростом n увеличивается также и N. На самом деле, N скорее всего будет расти гораздо быстрее n. Например, N может быть примерно пропорционально n^2, или n^3, или, скажем, 2^n (которое при больших n значительно превосходит n^2, n^3, n^4, n^5 и, вообще, n^r для любого фиксированного n), или даже 2^{2^n} (которое, в свою очередь, растет еще быстрее).

Конечно, число «шагов» зависит от типа вычислительной машины, на которой применяется алгоритм. Если эта машина принадлежит классу машин Тьюринга, описанному в главе 2, у которых есть только одна лента — что довольно неэффективно — то число N может расти еще быстрее (или, эквивалентно, машина будет работать медленнее), чем в случае с двумя и более лентами. Чтобы избежать этих неопределенностей, вводится широкая классификация всех возможных зависимостей $N(n)$, так что, независимо от типа используемой машины Тьюринга, величина темпов роста N будет всегда попадать в одну и ту же категорию. Одна из таких категорий, известная как **P** (от названия «полиномиальное время»), включает все темпы роста, которые являются фиксированными кратными n или $n^2, n^3, n^4, n^5, \ldots$ [12]. Это означает, что для любой задачи, попадающей в эту категорию **P** (под «задачей» здесь фактически понимается семейство задач, решаемых с помощью единого алгоритма), будет справедлива оценка

$$N \leqslant K \times n^r,$$

где K и r — константы, не зависящие от n. То есть N не может быть больше, чем число, кратное n в некоторой фиксированной степени.

Простой пример задачи, безусловно относящейся к **P**, — перемножение двух чисел. Чтобы объяснить это, я должен сначала описать, как число n характеризует размер двух чисел, которые надо перемножить. Мы можем принять, что оба числа представлены в двоичной записи и что $n/2$ — это просто количество бинарных разрядов в каждом из чисел, так что *общее число цифр* (то есть *битов*) у обоих равно n. (Если одно из чисел длиннее другого, то мы можем записать более короткое, начав с дополнительной последовательности нулей, тем самым выровняв их по длине.) Например, если $n = 14$,

[12] Понятие «полиномиальный» применяется на самом деле к выражениям более общего вида, скажем, $7n^4 - 3n^3 + 6n + 15$, но это никак не изменит общности наших рассуждений. В любом таком выражении все члены младших степеней n теряют значимость по мере увеличения n (так что в данном примере можно игнорировать все слагаемые, кроме $7n^4$).

мы бы могли рассмотреть произведение

$$1011010 \times 0011011,$$

которое является, на самом деле, произведением 1011010×11011, но с добавленными перед более коротким числом нулями. Выполнить требуемое действие проще всего путем умножения «в столбик»:

```
         1011010
        ×0011011
        --------
         1011010
        1011010
       0000000
      1011010
     1011010
    1000000
   0000000
   -------------
   0100101111110
```

учитывая, что в двоичной системе $0 \times 0 = 0$, $0 \times 1 = 0$, $1 \times 0 = 0$, $1 \times 1 = 1$, $0 + 0 = 0$, $0 + 1 = 1$, $1 + 0 = 1$, $1 + 1 = 10$. Число отдельных двоичных перемножений равно $(n/2) \times (n/2) = n^2/4$, а число отдельных двоичных сложений может доходить до $n^2/4 - n/2$ (включая перенос). Это дает $n^2/2 - n/2$ отдельных арифметических операций — и мы должны еще учесть несколько дополнительных логических шагов, которые задействованы в операциях переноса. Тогда общее число шагов, игнорируя члены более низкого порядка, равно по существу $N = n^2/2$, что, очевидно, является полиномом [13].

В общем случае, мы полагаем «размер» n задачи из некоторого класса равным полному количеству двоичных цифр (или битов), необходимых для задания свободных входных данных в задаче указанного размера. Другими словами, для произвольного размера n задача может иметь до 2^n различных вариантов (ибо для каждой из цифр имеется две возможности — 0 или 1, — а общее количество цифр равно n), и все они должны одинаково обрабатываться алгоритмом не более, чем за N шагов.

Существует масса примеров (классов) задач, которые не «принадлежат» множеству **P**. Например, чтобы вычислить 2^{2^r} для заданного натурального r, нам только *для записи конечного ответа* потребуется около 2^n шагов (где n — число цифр в двоичной записи r), не говоря даже о самом вычислении. Операция по вычислению $2^{2^{2^r}}$ потребует уже 2^{2^r} шагов для записи и так далее. Значения этих выражений намного превосходят те, которые дают полиномы для тех же n, и, следовательно, не могут принадлежать **P**.

Больший интерес представляют задачи, в которых ответ может быть записан и даже проверен на верность за «полиномиальное» время. Есть очень важная категория (алгоритмически решаемых классов) проблем, обладающих таким свойством. Их называют **NP**-задачами (классом задач). Точнее, если некоторая задача из класса **NP** имеет решение, то алгоритм позволит получить это решение, которое затем может быть проверено за «полиномиальное» время. Если же задача не имеет решения, то алгоритм сообщит об этом, но при этом не оговаривается необходимость проверки этого факта за «полиномиальное» или какое бы то ни было время [14].

NP-задачи встречаются во многих областях, причем как в математике, так и в повседневной практике. Я приведу здесь только один простой математический пример: задачу нахождения так называемого «*гамильтонова цикла*» на графе (довольно устрашающее название для чрезвычайно простой идеи). Под графом подразумевается конечный набор точек, или «вершин», некоторое количество пар которых соединено между собой линиями — «сторонами» графа. (Нас не интересуют сейчас геометрические или линейные свойства, а только то, какие вершины соединяются друг с другом. Поэтому не имеет значения, лежат ли все вершины в одной плоскости — если нас не волнует возможность пересечения двух сторон — или же в трехмерном пространстве.) Гамильтонов цикл — это замкнутый маршрут (петля), состоящий только из сторон графа и проходящий не более одного раза через любую из вершин. Пример графа с изображенным на нем гамильтоновым циклом показан на рис. 4.14. Задача нахождения гамильтонова цикла заключается в том, чтобы определить, существует ли гамильтонов цикл на рассматриваемом графе, и если существует, то явным образом указать его.

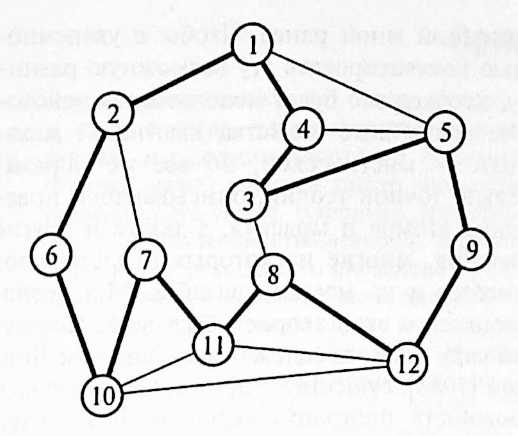


Рис. 4.14. Граф с гамильтоновым циклом (изображен зачерненными линиями). Существует только один гамильтонов цикл, как читатель может сам убедиться

Есть разные способы представления графов на языке двоичных чисел. Неважно, какой из этих способов применяется в том или ином случае. Один из методов заключается в том, чтобы пронумеровать вершины 1, 2, 3, 4, 5, ..., а потом перечислить пары в некотором подходящем фиксированном порядке:

$$(1,2),(1,3),(2,3),(1,4),(2,4),(3,4),(1,5),\\(2,5),(3,5),(4,5),(1,6),\ldots.$$

Затем мы на место каждой пары помещаем «1», если пара соединена стороной графа, и «0» — в противном случае. Тогда двоичная последовательность

$$10010110110\ldots$$

будет означать, что вершина 1 соединяется с вершинами 2, 4 и 5; вершина 3 — с вершинами 4 и 5; вершина 4 — с вершиной 5, и т. д. (в соответствии с рис. 4.14). Гамильтонов цикл может быть задан по желанию просто как подмножество этих сторон, которое было бы описано такой же двоичной последовательностью, как и ранее, но со значительно бо́льшим числом нулей. Процедура проверки в этом случае проходит несравненно быстрее, чем процесс непосредственного построения гамильтонова цикла. Все, что нужно выяснить, — это является ли построенный цикл действительно циклом, т. е. принадлежат ли его стороны исходному графу, и что каждая вершина графа используется ровно два раза — по одному разу на концах каждой из входящих в нее двух сторон. Такую процедуру проверки можно легко завершить за «полиномиальное» время.

На самом деле эта задача относится не только к **NP**, но к так называемой категории **NP**-*полных* задач. Это означает, что любая другая **NP**-задача может быть сведена к данной за «полиномиальное» время — так что, если бы кому-нибудь удалось отыскать алгоритм для решения задачи нахождения гамильтонова цикла за «полиномиальное» время (т. е. показать, что задача гамильтонова цикла действительно принадлежит **P**), то это будет означать, что все **NP**-задачи будут лежать в **P**! Это имело бы очень важные следствия. В широком смысле, задачи из **P** считаются *«податливыми»* (иначе говоря, «решаемыми за приемлемое время») для относительно больших n, на быстром современном компьютере; тогда как задачи из **NP**, но не лежащие в **P**, считаются *«неподатливыми»* (т. е. решаемыми в принципе, но «нерешаемыми практически») для тех же n — независимо от того, на какое разумно предсказуемое увеличение быстродействия компьютеров рассчитывать в будущем. (Реальное время, которое бы потребовалось для достаточно больших n при решении «неподатливой» задачи, легко превосходит возраст вселенной, что никак не предполагает практическое использование такого подхода!) Любой «умный» алгоритм для решения задачи о нахождении гамильтонова цикла за «полиномиальное» время мог бы быть превращен в алгоритм для решения *всех* прочих **NP**-задач, и тоже за «полиномиальное» время!

Другая задача, также являющаяся **NP**-полной [15] — «задача коммивояжера», которая во многом похожа на гамильтонов цикл, если не считать того, что разным сторонам приписаны числа и ставится цель отыскать гамильтонов цикл с *минимальной* суммой этих чисел (минимальной «длиной» пути, проделанного коммивояжером). Аналогично, «полиномиальное» время решения, достигнутое в «задаче коммивояжера», привело бы к возможности решать все **NP**-задачи за «полиномиальное» время. (Если такое решение когда-нибудь найдется, то новость об этом сразу попала бы на первые страницы! Ведь к **NP**-задачам относится, в частности, факторизация больших

целых чисел, которая применяется в секретных шифровальных системах, представленных за последние несколько лет. Если эта задача окажется решаемой за «полиномиальное» время, то, возможно, такие шифры могли бы быть взломаны при помощи мощных современных компьютеров; если же нет — эти шифры останутся неприступными. См. Гарднер [1989].)

Эксперты, как правило, полагают, что используя устройство, работающее по принципу машины Тьюринга, невозможно за «полиномиальное» время решить **NP**-полную задачу; и что, следовательно, **P** и **NP** — неэквивалентны. Это мнение, похоже, верно, хотя пока его никто не смог доказать. И это остается наиболее важной и на сегодняшний день нерешенной задачей теории сложности.

Сложность и вычислимость в физических объектах

Теория сложности является важной для наших рассуждений в этой книге не только потому, что она касается вопроса возможности алгоритмизации, но и потому, что она позволяет для заведомо алгоритмизуемых объектов решать вопрос о том, могут ли использоваться соответствующие алгоритмы *на практике*. В последующих главах я буду больше говорить о вычислимости, чем о теории сложности, поскольку я склонен думать (хотя, конечно, и не имея для этого достаточных оснований), что, в отличие от фундаментального вопроса вычислимости, положения теории сложности не настолько значимы для феномена мышления. Более того, мне представляется, что теория сложности сегодня лишь слегка затрагивает вопросы практичности алгоритмов.

Однако, я могу кардинально ошибаться по поводу важности той роли, которую играет сложность. Как будет показано позднее (глава 9, с. 343), теория сложности для *реальных физических объектов*, вероятно, может существенно отличаться от теории, изложенной мной ранее. Чтобы с уверенностью констатировать эту возможную разницу, необходимо будет использовать некоторые волшебные свойства квантовой механики — мистической, но все же поразительно точной теории, описывающей поведение атомов и молекул, а также и другие явления, многие из которых представляют интерес и на макромасштабах. Мы познакомимся с этой теорией в главе 6. Согласно ряду идей, предложенных Давидом Дойчем [1985], существует принципиальная возможность построить «квантовый компьютер», на котором за «полиномиальное» время могут быть решены некоторые задачи (или классы задач), не принадлежащих P. Пока совершенно неясно, как на практике сконструировать такое физическое устройство, которое бы (надежно) функционировало по принципу «квантового компьютера» — и, более того, рассматриваемый до сих пор класс задач носил заведомо искусственный характер, — но *теоретически* понятно, что квантовое физическое устройство могло бы улучшить работу машины Тьюринга.

А есть ли вероятность, что человеческий мозг, который в рамках данного обсуждения я рассматриваю как физическое устройство, хотя и имеющее чрезвычайно тонкую и сложную структуру — может неким образом использовать волшебство квантовой теории? Понимаем ли мы сегодня, как именно квантовые эффекты могут с пользой применяться для решения задач или формирования суждений? Можем ли мы представить, что для использования этих возможных преимуществ нам придется выйти «за нынешние пределы» квантовой теории? Насколько вероятно усовершенствование реальных физических устройств с учетом теории сложности для машин Тьюринга? И что говорит о таких устройствах теория *вычислимости*?

Чтобы рассматривать эти вопросы, нам надо будет отойти на время от математических абстракций и задаться целью выяснить в следующих главах, как же, в действительности, ведет себя окружающий нас мир!

Примечания

1. Рассматривая множества, члены которых, в свою очередь, также являются множествами, мы должны тщательно проводить отличия между членами такого множества и членами его *членов*. Например, допустим, что S — это множество *непустых* подмножеств некоторого другого множества T, членами которого являются один апельсин и одно яблоко. T в таком случае имеет свойство «двойственности», тогда как S обладает свойством «тройственности»: членами S будут множества а) из одного яблока; б) из одного апельсина и в) из одного апельсина и одного яблока — представляющие *три* члена S. Аналогично, множество, чьим единственным членом является *пустое множество*, будет иметь свойство «единичности», а не «нулевости» — в него входит один член, а именно пустое множество! При этом *само* пустое множество не будет иметь, конечно, ни одного члена.

2. В действительности ход рассуждений в теореме Геделя может быть представлен таким образом, чтобы не зависеть от полностью привнесенного извне понятия «истины» для утверждений, подобных $P_k(k)$. Однако, он по-прежнему будет зависеть от интерпретации фактического «значения» *некоторых* символов: в частности, «$\sim \exists$ » должно *означать* «не существует (натурального числа) ...такого, что...».

3. В нижеследующем прописные буквы будут представлять натуральные числа, а заглавные — конечные множества натуральных чисел. Пусть $m \to [n, k, r]$ представляет такое утверждение: «Если $X = \{0, 1, \ldots, m\}$, каждое из подмножеств которого длиной в k элементов приписано к r ящикам, то существует „большое" подмножество Y, принадлежащее X и имеющее по крайней мере n элементов, такое, что все подмножества Y из k элементов попадут в один ящик». Здесь «большое» означает, что число элементов, входящих в Y, больше самого маленького из натуральных чисел, принадлежащих Y. Рассмотрим теперь следующее утверждение: «При любых k, r, n существует m_0 такое, что при $m > m_0$ утверждение $m \to [n, k, r]$ всегда справедливо». Дж. Парисом и Л. Харрингтоном [1977] было доказано, что это положение эквивалентно геделевскому утверждению для стандартных (введенных Пеано) аксиом арифметики, которое не выводится из этих аксиом и которое позволяет делать утверждения о тех аксиомах, которые «очевидно верны» (в данном случае оно говорит, например, о том, что утверждения, выведенные из аксиом, сами будут справедливыми).

4. Статья называлась «Система логики, основанная на порядковых числах», и некоторые читатели будут уже знакомы со способом записи Канторовых *порядковых чисел*, который я применял для субиндексов. Иерархия логических систем, которые получаются с помощью приведенной мной процедуры, описывается с помощью *вычислимых порядковых чисел*.

   Есть несколько довольно естественных и легко формулируемых математических теорем, которые, если их пытаться доказать путем использования стандартных (введенных Пеано) правил арифметики, привели бы к «гипертрофированной» геделевской процедуре (по числу шагов многократно превосходящей ту, что я описал ранее). Математические доказательства этих теорем по природе своей не зависят от туманных и сомнительных рассуждений, выходящих за рамки аппарата нормального математического доказательства (см. Сморински [1983]).

5. Континуум-гипотеза, которая упоминалась в главе 3, с. 101 (и из которой следует, что $C = \aleph_1$), является наиболее «экстремальным» математическим утверждением, которое здесь встречается (хотя часто рассматриваются и куда более «экстремальные» рассуждения). Континуум-гипотеза интересна еще и потому, что сам Гедель, совместно с Полом Дж. Коэном, показал, что эта гипотеза в действительности *не зависит* от стандартных аксиом и правил теории множеств. Таким образом, отношение любого математика к континуум-гипотезе позволяет причислить его к сторонникам либо формалистской, либо платонистской точки зрения. Для формалиста данная гипотеза будет «недоказуемой», поскольку ее справедливость не может быть установлена или опровергнута, если опираться на стандартную (построенную Цермело и Френкелем) формальную систему, и, значит, не «имеет смысла» называть ее ни «истинной», ни «ложной». Однако, для убежденного платониста эта гипотеза является либо истинной, либо ложной, хотя какой именно — это можно установить только путем рассуждений некоторого нового типа, идущих еще дальше, чем использование геделевских утверждений для формальной системы Цермело—Френкеля. (Коэн [1966] сам предложил принцип рефлек-

сии, который позволяет показать, что континуум-гипотеза — «с очевидностью ложна»!)

6. Живое и не слишком насыщенное техническими деталями изложение этой темы можно найти у Ракера [1984].

7. Сам Брауэр начал размышлять в этом направлении, в частности, потому, что очень придирчиво и болезненно относился к «неконструктивности» доказательства своей теоремы из области топологии, «теоремы Брауэра о неподвижной точке». Эта теорема утверждает, что, если вы возьмете круг — то есть окружность вместе со всеми точками внутри нее — и будете непрерывно двигать его внутри области, где он находился изначально, то найдется по крайней мере одна точка круга, — называемая неподвижной точкой, — которая окажется точно там же, откуда она начала движение. Не ясно, где именно располагается эта точка, и может ли их быть несколько — теорема говорит только о *существовании* такой точки. (Среди математических теорем существования, эта, на самом деле, носит довольно «конструктивный» характер. Примеры теорем существования более высокой степени неконструктивности — это теоремы, зависящие от так называемой «аксиомы выбора» или «леммы Цорна» (см. Коэн [1966], Ракер [1984]).) Трудность в случае Брауэра была аналогична той, что возникает в следующей задаче: найти точки, в которых f обращается в нуль, если известно, что f — действительная непрерывная функция действительной переменной, которая принимает как положительные, так и отрицательные значения. Стандартная процедура заключается в последовательном делении пополам отрезка, на котором функция меняет свой знак; но решение о том, какое именно промежуточное значение принимает функция (положительное, отрицательное или нулевое), может оказаться неконструктивным в том смысле, которого требует Брауэр.

8. Мы нумеруем множества $\{v, w, x, \ldots, z\}$, где v представляет функцию f в согласии с некоторой лексикографической схемой. Мы (рекурсивно) проверяем на каждом этапе справедливость равенства

$$f(w, x, \ldots, z) = 0$$

и оставляем утверждение

$$\exists\, w, x, \ldots, z[(w, x, \ldots, z) = 0]$$

только в том случае, если это равенство выполняется.

9. Недавно я узнал от Леоноры Блюм, что (заинтересовавшись моими комментариями в первом издании этой книги) она установила, что множество Мандельброта (и его дополнение) на самом деле являются, как я и предполагал, нерекурсивными в том смысле, который описан в десятом примечании.

10. Блюмом, Шубом и Смэйлом [1989] была разработана новая теория вычислимости для действительных функций от действительных переменных (в отличие от общепринятых функций натуральных чисел, принимающих натуральные значения), подробности которой я узнал лишь совсем недавно. Эта теория применима и к комплексным функциям, а кроме того, может сыграть заметную роль в упомянутых мной вопросах.

11. Эта конкретная задача называется (если быть более точным) «задачей со словами для полугрупп». Существуют также и другие разновидности этой задачи, в которых действуют несколько отличные правила. Но нас это сейчас волновать не должно.

12. Более того, Ханф [1974] и Майерс [1974] показали, что существует отдельное множество (из большого числа «плиток»), которое покрывает плоскость только *невычислимым* образом.

13. В действительности, путем применения некоторых тонких ходов, можно сократить число шагов до величины порядка

$$n \log n \log (\log n)$$

для больших n — которая, конечно, все еще принадлежит P. За подробностями я отсылаю читателя к Кнуту [1981].

14. Если быть точным, классы P, NP и NP-полный (см. с. 143) определены только для задач типа «да или нет» (скажем, когда заданы a, b, c и спрашивается, выполняется ли для них $a \times b = c$); но описания, приведенные в тексте, вполне подходят для наших целей.

15. Строго говоря, нам нужно переформулировать эту задачу под ответ «да или нет», например: существует ли маршрут для коммивояжера, длина которого меньше чем столько-то? (См. предыдущее примечание.)

Глава 5

Классический мир

Состояние физической теории

Что нам нужно знать о законах природы, чтобы понять, какая роль в ней может быть отведена сознанию? Насколько важно представлять себе для этого принципы организации и взаимодействия, которым подчиняются элементы, составляющие тело и мозг? Если осознанное восприятие — всего лишь результат выполнения алгоритмов (как нас пытаются убедить многие приверженцы ИИ), то вопрос о конкретном виде и действии этих принципов не имеет особого значения. Любое устройство, *способное* просчитать алгоритм, будет ничуть не хуже любого другого. Но, быть может, наше чувство осознания не сводится полностью к работе алгоритмов. И возможно, что детальное знание нашего внутреннего устройства и точных физических законов, управляющих той субстанцией, из которой мы состоим, может оказаться достаточно важным. Вероятно, нам понадобиться понять те фундаментальные физические свойства, которые лежат в основе самой природы вещества и определяют его поведение. Сегодня физика не достигла пока такого уровня, и ей предстоит еще раскрыть множество тайн и испытать немало глубоких озарений. Тем не менее большинство физиков и физиологов склонны считать, что мы *уже сейчас* располагаем достаточным знанием тех физических законов, которые управляют работой такого объекта средних размеров, как наш мозг. Хотя никто не оспаривает исключительную сложность головного мозга человека как физической системы и не отрицает существования значительного числа пробелов в наших знаниях о его детальной структуре и принципах работы, — все же лишь немногие осмелились бы утверждать, что мы испытываем существенную нехватку знаний именно в области *физических* основ функционирования мозга.

Ниже я приведу пример, свидетельствующий как раз об обратном, — то есть о том, что мы еще *не знаем* физику настолько, чтобы (даже в принципе) иметь возможность адекватно использовать ее язык для описания работы человеческого мозга. Но прежде мне потребуется дать хотя бы в общих чертах представление о достижениях и состоянии современной физической теории. В этой главе речь пойдет главным образом о той области, которую принято называть «классической физикой» и которая включает в себя механику Ньютона и теорию относительности Эйнштейна. По существу, термин «классическая» в данном случае означает, что обе теории достигли расцвета задолго до рождения (примерно в 1925 году, вдохновенными трудами таких физиков, как Планк, Эйнштейн, Бор, Гейзенберг, Шредингер, де Бройль, Борн, Иордан, Паули и Дирак) *квантовой теории* — загадочной теории, опирающейся на вероятности и индетерминизм и описывающей поведение молекул, атомов и субатомных частиц. В отличие от квантовой теории, классическая теория является *детерминистской*, поэтому

будущее в ее рамках всегда полностью определяется прошлым. Но даже и в классической физике есть еще много загадок, несмотря на то, что знания, накопленные за несколько веков, позволили нам построить феноменально точную картину мира. Мы также должны будем рассмотреть и квантовую теорию (в главе 6), ибо я убежден, что — несмотря на мнение, разделяемое большинством физиологов — квантовые явления *могут* играть важную роль в функционировании головного мозга человека. Но к этой теме мы обратимся в последующих главах.

К сегодняшнему дню наука достигла поразительных успехов. Достаточно бросить хотя бы беглый взгляд вокруг, чтобы воочию убедиться в невероятном могуществе, которое мы обрели благодаря нашему пониманию законов природы. Конечно, при создании современных технологий существенно использовались обширнейшие эмпирические данные. Однако куда более важна физическая *теория*, лежащая в основе этих технологий, и сейчас нас будет интересовать именно она. Теории, существующие в настоящее время, отличаются удивительной точностью. Но сила их заключается не только в способности правильно описывать соответствующие явления и процессы. Не меньшее значение имеет и то, что они, как оказывается, прекрасно поддаются точному и скрупулезному математическому анализу. Взятые вместе, эти обстоятельства позволили нам создать науку, обладающую поистине впечатляющей силой.

Физическая теория, о которой идет речь, имеет богатую историю. Но одно событие можно особенно выделить: это публикация в 1687 году *Математических начал натуральной философии* Исаака Ньютона. В этой работе, имеющей непреходящее значение, было показано, как, исходя из весьма немногих физических принципов, можно понять (причем зачастую с поразительной точностью) реальное поведение многих физических объектов. (Значительная часть *Начал* посвящена разработке математических методов, хотя более удобный для практического использования аппарат был создан позднее Эйлером и другими физиками и математиками.) Собственные труды Ньютона, как он охотно признавал, во многом опирались на труды его предшественников, выдающихся мыслителей, среди которых были Галилео Галилей, Рене Декарт и Иоганн Кеплер. Некоторые из основополагающих идей Ньютон заимствовал у еще более древних мыслителей. Упомяну в частности геометрические идеи Платона, Евдокса, Евклида, Архимеда и Аполлония. Об этих математиках я еще расскажу более подробно в дальнейшем.

Отклонения от основных положений динамики Ньютона появились позднее. Первым из них оказалась электромагнитная теория Джеймса Клерка Максвелла, разработанная в середине XIX века. Она охватывала не только классическое поведение электрического и магнитного полей, но и поведение света [1]. Эта замечательная теория будет рассмотрена нами чуть позднее. Теория Максвелла имеет первостепенное значение для современной технологии, равно как и для понимания принципов функционирования нашего головного мозга, в котором электромагнитные явления играют очень важную роль. Менее ясно, имеют ли какое-нибудь отношение к процессам нашего мышления две поистине великие теории относительности, связанные с именем Альберта Эйнштейна. *Специальная* теория относительности, возникшая из исследований уравнений Максвелла, была создана Анри Пуанкаре, Хендриком Лоренцем и Эйнштейном (позднее элегантное геометрическое описание специальной теории относительности предложил Герман Минковский) для объяснения необычного поведения тел, движущихся со скоростями, близкими к скорости света. Частью этой теории стало знаменитое соотношение Эйнштейна $E = mc^2$. Но влияние специальной теории относительности на технологию до сих пор остается весьма слабым (если не считать ядерной физики), а отношение к функционированию нашего мозга — в лучшем случае косвенным. С другой стороны, специальная теория относительности затрагивает фундаментальные вопросы физической реальности, связанные с природой *времени*. В последующих главах мы увидим, что это приводит нас к ряду «загадок» из области квантовой теории, которая может иметь принципиальное значение для

понимания наших механизмов восприятия «течения времени». Кроме того, нам необходимо понять специальную теорию относительности прежде, чем мы сможем должным образом оценить *общую* теорию относительности Эйнштейна — теорию, которая использует для описания гравитации искривленное пространство-время. До сих пор эта теория не оказывала на технологию почти никакого влияния [1], так что предположение о возможной связи между общей теорией относительности и процессами, происходящими в нашем мозге, потребовало бы немалой смелости воображения.

Интересно, что в наших дальнейших размышлениях *общая* теория относительности будет играть существенную роль, особенно в главах 7 и 8, где нам придется отправиться в самые удаленные области пространства и времени, чтобы собрать «по зернышку» сведения о тех изменениях, которые, как я считаю, необходимы для создания полностью непротиворечивой картины квантовой теории — но об этом позже!

Все, что мы до сих пор упоминали, относится к области *классической* физики. А как обстоит дело с квантовой физикой? В отличие от теории относительности, квантовая теория *начинает* оказывать существенное влияние на технологию. Отчасти это объясняется тем вкладом, который квантовая теория внесла в столь технологически важные области, как химия и металлургия. Действительно, для многих эти области теперь срослись с физикой именно благодаря тем новым знаниям, которые дала нам квантовая теория. Помимо этого существуют и совершенно *новые* явления, появление которых без квантовой теории было бы невозможным (самым известным из таких явлений, думаю, будет справедливо назвать лазер). Тогда что же мешает нам предположить, что некоторые существенные аспекты квантовой теории могут играть решающую роль в той физике, которая лежит в основе наших процессов мышления?!

А как обстоит дело с относительно новыми физическими теориями? Возможно, некоторым читателям приходилось встречаться с возбуждающими идеями, использующими такие понятия, как «кварки» (см. с. 151), теории великого объединения, «инфляционный сценарий» (см. заключительное примечание 13 на с. 302), «суперсимметрия», теория «(супер)струн» и т. д. Как такие новые течения согласуются с теми теориями, о которых шла речь выше? Насколько нам важно уделять внимание их изучению? Чтобы выработать ясное понимание подобных вопросов, было бы полезно разбить основные физические теории на три широкие категории. Я назову их следующим образом:

1. ПРЕВОСХОДНАЯ.
2. ПОЛЕЗНАЯ.
3. ПРОБНАЯ.

К категории ПРЕВОСХОДНАЯ надлежит отнести все теории, которые я рассматривал в предыдущих разделах. Для того, чтобы теорию можно было причислить к разряду ПРЕВОСХОДНЫХ, совершенно не обязательно, по-моему, требовать от нее полного согласия со всеми явлениями в мире — однако диапазон явлений и точность их описания должны быть в определенном смысле *феноменальными* [2]. Принимая во внимание такую трактовку термина «превосходная», остается только удивляться тому, что в эту категорию вообще попадают какие-то теории! Я не знаю ни одной фундаментальной теории в любой другой естественной науке, которую можно было бы с достаточным основанием отнести к этой категории. Возможно, больше всех других название «превосходной» заслуживает теория естественного отбора, выдвинутая Дарвином и Уоллисом, — но и ей далеко до идеала.

Самой древней из ПРЕВОСХОДНЫХ теорий по праву можно считать евклидову геометрию, с отдельными положениями которой мы познакомились еще в школе. Возможно, древние вообще не рассматривали

[1] Но только почти: точность, которая требуется при управлении полетом спутника или космического зонда, такова, что для расчета его орбиты нужно принимать во внимание эффекты искривления пространства-времени, описываемые общей теорией относительности. И то же самое справедливо в отношении особо прецизионных устройств, способных установить местоположение заданного объекта на земной поверхности с точностью до долей метра.

[2] В оригинале игра слов: *phenomen* — явление, *phenomenal* — феноменальный. — *Прим. ред.*

евклидову геометрию как физическую теорию, но в действительности она была таковой: тонкой и в высшей степени точной теорией физического пространства — и геометрии твердых тел. Почему я упоминаю о евклидовой геометрии как о *физической* теории, а не как о разделе математики? Причина проста: по иронии судьбы евклидова геометрия — как нам стало теперь известно — *не вполне точна* в качестве инструмента для описания того физического пространства, в котором мы все обитаем! Общая теория относительности Эйнштейна говорит нам, что пространство-(время) в действительности «искривлено» (т. е. *не является* в точности евклидовым) в присутствии гравитационного поля. Но этот факт отнюдь не лишает евклидову геометрию права называться ПРЕВОСХОДНОЙ теорией. Действительно, в метровом диапазоне отклонения от евклидовой плоскостности чрезвычайно малы, и ошибки, связанные с заменой геометрии реального пространства на евклидову, составляют величину меньшую, чем диаметр атома водорода!

С полным основанием можно утверждать, что *статика* — теория, занимающаяся изучением неподвижных тел — превратившаяся в красивую науку благодаря Архимеду, Паппу и С. Стевину — может быть смело отнесена к категории ПРЕВОСХОДНЫХ теорий. В настоящее время статика входит в ньютоновскую механику. Глубокие идеи *динамики* (занимающейся изучением движущихся тел) были заложены примерно в 1600 году Галилеем и позднее превращены Ньютоном в величественную и широкую по своему охвату теорию. Динамика несомненно должна быть включена в категорию ПРЕВОСХОДНЫХ теорий. Применительно к движению планет и лун экспериментальная точность динамики поистине превосходна — выше одной десятимиллионной. Одна и та же ньютоновская схема применима и здесь, на Земле, и за пределами звезд и галактики, причем примерно с одинаковой точностью. Аналогичным образом, теория Максвелла применима с высокой точностью в необычайно широком диапазоне, примыкающем с одного конца к микроскопическим масштабам атомов и субатомных частиц, а с другого — к масштабам галактик, т. е. в миллион миллионов миллионов миллионов миллионов миллионов раз больших! (На микроскопическом конце шкалы уравнения Максвелла необходимо надлежащим образом сочетать с правилами квантовой механики.) Так что теорию Максвелла по праву можно тоже отнести к ПРЕВОСХОДНЫМ теориям.

Специальная теория относительности Эйнштейна (предтечей которой выступил Пуанкаре, а изящную формулировку предложил Минковский) дает удивительно точное описание явлений, в которых скорости объектов могут приближаться к скорости света, т. е. при таких скоростях, когда ньютоновские описания начинают «не срабатывать». Изящная и оригинальная теория общей относительности Эйнштейна обобщает динамическую теорию (гравитации) Ньютона и повышает ее точность, наследуя при этом все достоинства теории Ньютона во всем, что касается движения планет и лун. Кроме того, общая теория относительности Эйнштейна объясняет различные необычные наблюдаемые явления, не укладывающиеся в более старую ньютоновскую схему. Рассмотрение одного из таких явлений (а именно, «двойного пульсара», см. с. 195) показывает, что теория Эйнштейна справедлива с точностью до 10^{-14}. Обе теории относительности, вторая из которых включает в себя первую, с полным основанием могут быть отнесены к категории ПРЕВОСХОДНЫХ теорий (по причинам их математического изящества, почти не уступающего их точности).

Диапазон явлений, объясняемых необычайно красивой и революционной квантовой механикой, и точность, с которой она согласуется с экспериментом, ясно указывают на то, что квантовая теория вне всяких сомнений может быть отнесена к категории ПРЕВОСХОДНЫХ. Никаких расхождений между наблюдениями и квантовой механикой не известно — но сила ее простирается еще дальше, проявляя себя в ряде ранее необъяснимых явлений, которые ныне получили обоснование в рамках этой теории. Законы химии, стабильность атомов, четкость спектральных линий (см. с. 210) и их весьма специфическое расположение в наблюдаемых спектрах; удивительное явление

сверхпроводимости (нулевого электрического сопротивления) и поведение лазеров — таков далеко не полный перечень явлений, объясняемых квантовой механикой.

Я устанавливаю высокие стандарты для категории ПРЕВОСХОДНЫХ теорий — но именно к таким стандартам мы привыкли в физике. А как обстоит дело с теориями, появившимися в последнее время? По моему мнению, только одна из них может претендовать на включение в категорию ПРЕВОСХОДНЫХ, и она не так уж нова: я имею в виду теорию, получившую название *квантовой электродинамики* (или КЭД). Ее основы заложили в своих трудах Иордан, Гейзенберг и Паули; сформулирована она была Дираком в 1926–1934 годах; а «рабочую форму» обрела в работах Бете, Фейнмана, Швингера и Томонаги в 1947–1948 годах. Эта теория возникла как соединение принципов квантовой механики и специальной теории относительности, совместно с уравнениями Максвелла и фундаментальным уравнением, описывающим движение и спин электронов, выведенным Дираком. В целом, квантовая электродинамика не обладает привлекательным изяществом или непротиворечивостью более ранних ПРЕВОСХОДНЫХ теорий, но я, тем не менее, считаю возможным отнести эту дисциплину к таковым в силу ее поистине феноменальной точности. Особого упоминания заслуживает хотя бы один результат, следующий из квантовой электродинамики — оценка величины магнитного момента электрона. (Электроны ведут себя как крохотные магниты, образованные вращающимися вокруг собственной оси электрическим зарядом. Термин «магнитный момент» как раз и характеризует силу такого крохотного магнита.) Величина 1,00115965246 (в соответствующих единицах и с допустимой погрешностью около 20 в двух последних знаках) была вычислена для магнитного момента электрона на основе квантовой электродинамики — в то время как самое последнее из полученных экспериментальных значений этой величины составляет 1,001159652193 (с возможной погрешностью около 10 в двух последних цифрах). Как отметил Фейнман, при столь малой погрешности расстояние от Нью-Йорка до Лос-Анджелеса можно было бы определить с точностью до толщины человеческого волоса! Нам нет необходимости досконально знакомиться здесь с этой теорией, но для создания у читателя более полного представления о предмете наших рассуждений, я в конце следующей главы вкратце упомяну некоторые из принципов и существенных особенностей квантовой электродинамики [3].

Отдельные современные теории я мог бы отнести к категории ПОЛЕЗНЫХ. Две из них не понадобятся нам в дальнейшем, но упомянуть о них все же стоит. Первая — это *кварковая* модель субатомных частиц Гелл-Манна—Цвейга. Субатомные частицы называются *адронами*. К этой группе относятся протоны, нейтроны, мезоны и т. д., образующие атомные ядра, — или, точнее, «сильно взаимодействующие» частицы. Возникшая (позднее) детальная теория их взаимодействия получила название *квантовой хромодинамики*, или КХД. Основная идея КХД состоит в том, что все адроны «построены» из составных частей, называемых «кварками», которые взаимодействуют между собой в соответствии с некоторым обобщением теории Максвелла (известным под названием «теории Янга—Миллса»). Во-вторых, существует теория (предложенная Глэшоу, Саламом, Уордом и Вайнбергом — также на основе теории Янга—Миллса), объединяющая электромагнитное взаимодействие со «слабым» взаимодействием, ответственным за радиоактивный распад. Эта теория включает в себя описание так называемых *лептонов* (электронов, мюонов, нейтрино, а также W- и Z-частиц — т. е. всех «слабо взаимодействующих» частиц). Обе теории подкрепляются солидными экспериментальными данными. Но по различным причинам эти теории не столь точно, как хотелось бы (по сравнению, например, с КЭД или другими теориями), согласуются с экспериментом, и их предсказательная сила в настоящее время еще далеко не соответствует тем *феноменальным* стандартам, которые требуются для их включения в категорию ПРЕВОСХОДНЫХ теорий. Взятые вместе, эти две теории (причем вторая

[3] Популярное изложение квантовой электродинамики см. в книге Фейнмана *КЭД* [1985].

из них — вместе с КЭД) иногда называются *стандартной моделью*.

Наконец, существует еще одна теория (другого типа), которая, на мой взгляд, относится по меньшей мере к категории ПОЛЕЗНЫХ теорий. Я говорю о теории *Большого взрыва*, в результате которого родилась Вселенная[4)]. Эта теория будет играть важную роль в главах 7 и 8.

На этом, как мне кажется, заканчивается список теорий — претендентов на звание ПОЛЕЗНЫХ [2]. Существует много идей, пользующихся в настоящее время (или пользовавшихся до недавнего времени) широкой популярностью. Среди них «теории Калуцы—Клейна»; «суперсимметрия» (или «супергравитация»); все еще чрезвычайно модные теории «струн» (или «суперструн»); а также *теории великого объединения*, равно как и отдельные порожденные ими идеи, например, «инфляционный сценарий» (см. примечание 13 на с. 302). Все они, по моему твердому убеждению, относятся к категории ПРОБНЫХ теорий (см. работы Барроу [1988], Клоса [1983], Дэвиса и Брауна [1988], Сквайерса [1985]). Важное различие между категориями ПОЛЕЗНЫХ и ПРОБНЫХ теорий состоит в том, что последние не подкреплены надежными экспериментальными данными [3]. Это отнюдь не означает, что какая-нибудь из них не может неожиданно возвыситься до разряда ПОЛЕЗНОЙ и даже ПРЕВОСХОДНОЙ. Некоторые из упомянутых выше теорий содержат оригинальные и весьма многообещающие идеи, пока, правда, не получившие достаточного экспериментального подтверждения. Категория ПРОБНЫХ теорий охватывает весьма широкий диапазон. Не исключено, что концепции, встречающиеся в отдельных теориях подобного рода, несут в себе зерна новых достижений в понимании природы — но в то же время другие из них на удивление неправдоподобны и вполне могут ввести своих сторонников в заблуждение. (У меня было искушение отщепить от категории почтенных ПРОБНЫХ теорий еще одну, четвертую категорию, и назвать ее, скажем, ТУПИКОВЫЕ теории, но по зрелом размышлении я отказался от этого намерения, поскольку не хочу потерять половину своих друзей!)

Не следует удивляться тому, что основные ПРЕВОСХОДНЫЕ теории возникли довольно давно. Вероятно, на протяжении истории таких теорий существовало гораздо больше, но некоторые из них со временем перешли в категорию ПРОБНЫХ и в большинстве своем оказались забыты. Аналогичным образом, в категорию ПОЛЕЗНЫХ теорий попадало немало таких, которые впоследствии теряли свою актуальность, тогда как некоторые поглощались другими — ставшими впоследствии ПРЕВОСХОДНЫМИ теориями. Рассмотрим несколько примеров. До того, как Коперник, Кеплер и Ньютон создали новую, более совершенную теорию, существовала детально разработанная теория планетных движений, родившаяся в Древней Греции и получившая название *птолемеевой системы*. Согласно этой модели, движения планет описывались сложной суперпозицией круговых движений. Птолемеева система была весьма эффективной с точки зрения предсказаний, но с каждым разом становилась все сложнее и сложнее по мере повышения требований к точности. Нам, живущим ныне, птолемеева система кажется слишком искусственной. Это — хороший пример ПОЛЕЗНОЙ системы (она действительно была полезной на протяжении почти двадцати веков!), которая впоследствии, сыграв свою историческую организующую роль, *сошла со сцены* как физическая теория. В качестве хорошего примера ПОЛЕЗНОЙ теории, которая в конце концов доказала свою состоятельность, можно привести блестящую идею Кеплера о движении планет по эллиптическим орбитам. Другим примером могла бы стать периодическая система химических элементов Менделеева. Сами по себе эти идеи не позволяют построить модели, обладающие предсказательной силой требуемого «феноменального» характера, однако в будущем они становятся «правильными» следствиями из выросших из них ПРЕВОС-

[4)] Имеется в виду так называемая «стандартная модель» Большого взрыва. Существует много вариантов теории Большого взрыва, наиболее известный из которых носит имя «инфляционный сценарий». По моему твердому убеждению, инфляционный сценарий относится к категории ПРОБНЫХ теорий.

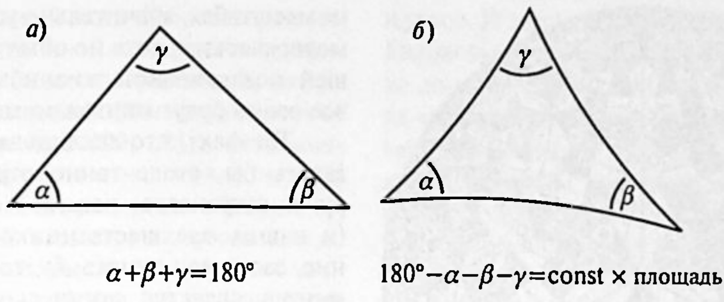


Рис. 5.1. *а)* Треугольник в евклидовом пространстве. *б)* Треугольник в пространстве Лобачевского

ХОДНЫХ теорий (соответственно, ньютоновской динамики и квантовой теории).

В последующих разделах и главах я не буду останавливаться на обсуждении существующих ныне теориях, которые всего лишь ПОЛЕЗНЫ или ПРОБНЫ. Достаточно сказать о тех теориях, которые ПРЕВОСХОДНЫ. Можно считать удачей, что у нас есть такие теории, позволяющие постигать этот мир во всей его полноте. Но в конечном счете, мы должны попытаться решить вопрос о том, достаточно ли могущественны даже эти теории, чтобы описывать функционирование нашего мозга и работу разума. В свое время я еще вернусь к этой теме — а пока мы рассмотрим ПРЕВОСХОДНЫЕ теории в том виде, в котором они нам сегодня известны, и попробуем оценить степень их применимости к интересующим нас задачам.

Евклидова геометрия

Евклидова геометрия — это, попросту говоря, тот самый предмет, который мы изучаем в школе как «геометрию». Однако я подозреваю, что большинство людей склонны считать евклидову геометрию областью математики, а вовсе не физической теорией. Разумеется, евклидова геометрия является *в том числе* и математикой — но все же это не единственная возможная математическая геометрия. Та геометрия, которую придумал Евклид, очень точно описывает физическое пространство нашего с вами мира, но это — не логически необходимое следствие, а всего лишь (почти точно) наблюдаемое свойство физического мира.

Действительно, существует другая геометрия, называемая геометрией *Лобачевского* (или *гиперболической*)[5], которая во многом похожа на евклидову геометрию, но имеет при этом и некоторые интригующие отличия. Напомним, в частности, что в евклидовой геометрии сумма углов треугольника всегда равна 180°. В геометрии Лобачевского сумма углов треугольника всегда *меньше* 180°, причем отличие суммы углов от 180° пропорционально площади треугольника (рис. 5.1).

Замечательный голландский художник Мориц К. Эшер создал несколько мозаик, очень тонко и точно передающих суть геометрии Лобачевского. Одна из этих мозаик представлена на рис. 5.2. Каждую черную рыбу, в соответствии с геометрией Лобачевского, следует считать имеющей такой же размер и такую же форму, что и любая другая черная рыба. Для белых рыб — аналогично. Геометрия Лобачевского не может быть абсолютно точно воспроизведена на евклидовой плоскости, отсюда — кажущееся скопление рыб вблизи круговой границы. Представьте себе, что вы находитесь внутри мозаики где-то у этой окружности. Тогда геометрия Лобачевского должна для вас выглядеть точно такой же, как если бы находились в центре или в каком-то другом месте мозаики. То, что выглядит как «граница» мозаики в этом евклидовом представлении, в действительности

[5] Николай Иванович Лобачевский (1792–1856) — один из нескольких математиков, независимо друг от друга открывших этот тип геометрии (альтернативный геометрии Евклида). Имена остальных: Карл Фридрих Гаусс (1777–1855), Фердинанд Швейкард и Янош Бойяи.

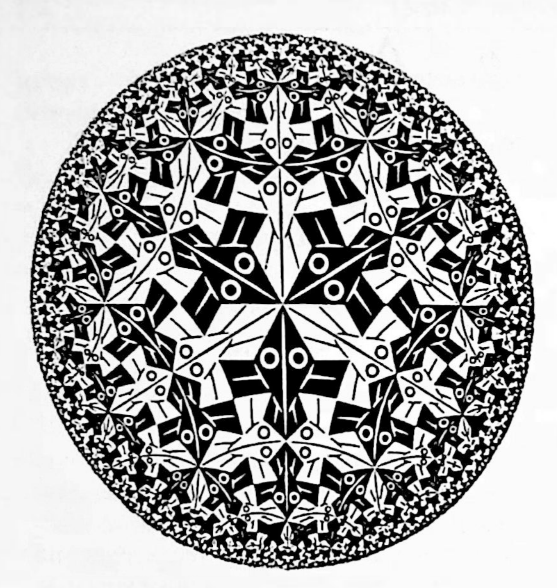


Рис. 5.2. Пространство Лобачевского, изображенное Эшером в виде мозаики. (Все рыбы — как черные, так и белые — должны считаться конгруэнтными.)

находится «на бесконечности» в геометрии Лобачевского. Граничную окружность вообще не следует рассматривать как часть пространства Лобачевского — равно как и никакую часть евклидовой области, лежащую *за* ее пределами. (Это остроумное представление плоскости Лобачевского принадлежит Пуанкаре. Его достоинство заключается в том, что форма очень маленьких фигур при этом не искажается — изменяются только их размеры.) «Прямыми» в геометрии Лобачевского (вдоль которых расположены некоторые из рыб на мозаике Эшера) служат окружности, пересекающие круговую границу под прямыми углами.

Вполне может быть, что геометрия Лобачевского действительно выполняется для нашего мира в космологических масштабах (см. главу 7, с. 284). Но коэффициент пропорциональности между дефицитом углов и площадью треугольника в этом случае *чрезвычайно* мал, а для обычных масштабов евклидова геометрия дает превосходное приближение геометрии Лобачевского. В самом деле, как мы увидим далее в этой главе, общая теория относительности Эйнштейна говорит нам о том, что геометрия нашего мира *действительно* отклоняется от евклидовой геометрии (хотя и «нерегулярно», т. е. более сложно, чем геометрия Лобачевского) на масштабах, значительно уступающих космологическим, хотя по обычным меркам нашей повседневной жизни эти отклонения все равно будут ничтожно малы.

Тот факт, что евклидова геометрия, казалось бы, столь точно отражает структуру «пространства» нашего мира, вводил нас (и наших предшественников!) в заблуждение, заставляя думать, будто евклидова геометрия является логической необходимостью или будто мы обладаем внутренней интуитивной способностью *априори* догадаться, что евклидова геометрия *должна* быть применима к миру, в котором мы живем. (Так утверждал даже великий философ Иммануил Кант.) Реальный разрыв с евклидовой геометрией наступил только с созданием Эйнштейном общей теории относительности, появившейся на свет много лет спустя. И тогда стало понятно, что евклидова геометрия вовсе не является логической необходимостью, и что ее весьма точное (хотя и далеко не абсолютное) соответствие структуре нашего физического пространства — не более, чем *результат эмпирических наблюдений*! Евклидова геометрия действительно была (ПРЕВОСХОДНОЙ) *физической* теорией. И это в дополнение к тому, что евклидова геометрия — изящный и логически непротиворечивый раздел чистой математики.

Здесь угадывается определенное сходство с философской концепцией Платона (изложенной примерно в 360 году до н. э. — почти за пятьдесят лет до появления *Начал* Евклида — знаменитого сочинения по геометрии). С точки зрения Платона объекты чистой геометрии — прямые, окружности, треугольники, плоскости и т. п. — могут быть лишь приблизительно реализованы в реальном мире физических вещей. Эти математически точные объекты чистой геометрии обитают в другом мире — *платоновском идеальном мире* математических понятий. Платоновский мир состоит не из осязаемых вещей, а из «математических объектов». Этот мир доступен нашему восприятию не обычным физическим путем, а посредством *интеллекта*. Человеческий разум контактирует с миром Платона всякий раз, когда открывает математическую истину, постигая ее с помощью математических рассуждений и интуитивных

догадок. Идеальный мир Платона рассматривался как отличный от нашего материального мира — более совершенный, но при этом столь же реальный. (Вспомним сказанное в главах 3 и 4, с. 109, 121 о платоновской реальности математических понятий.) Таким образом, хотя идеальные объекты чистой евклидовой геометрии можно исследовать с помощью мысли, логически выводя при этом их свойства — отсюда вовсе не следует, что для «несовершенного» физического мира, воспринимаемого нашими органами чувств, неукоснительное следование этому идеалу является необходимостью. Располагая в свое время достаточно скудными данными, Платон, по-видимому благодаря какому-то чудесному озарению, смог предугадать, что, с одной стороны, математику следует изучать и понимать ради самой математики, и что нельзя требовать полного и точного соответствия математических объектов объектам физического опыта; а с другой — что функционирование реального внешнего мира в конечном счете может быть понято только в терминах точной математики, т. е. в терминах платоновского идеального мира, «доступного через интеллект»!

Платоном в Афинах была основана Академия, в задачи которой входило дальнейшее развитие таких идей. Среди элиты, выросшей из числа членов этой Академии, был и необычайно влиятельный и знаменитый философ Аристотель. Но здесь нас будет интересовать другой человек, принадлежащий к платоновской Академии — математик и астроном Евдокс, несколько менее известный, чем Аристотель, но, по моему глубокому убеждению, гораздо более проницательный ученый, один из величайших мыслителей античности.

В евклидовой геометрии есть одна очень важная и тонкая составляющая, которая, на самом деле, является очень существенной и которую сегодня мы вряд ли вообще отнесли бы к геометрии! (Математики охотнее назвали бы это «анализом», чем «геометрией».) Речь идет о введении *действительных чисел*. Евклидова геометрия использует длины и углы. Чтобы иметь возможность использовать такую геометрию, нам необходимо понимать, какого рода «числа» нужны для описания этих самых длин и углов. И здесь новая идея была предложена Евдоксом (ок. 408–335 гг. до н. э.) в IV веке до н. э.[6] Греческая геометрия переживала «кризис» из-за открытия пифагорейцами таких чисел, как $\sqrt{2}$ (последнее необходимо для того, чтобы выразить длину диагонали квадрата через длины его сторон), не представимых в виде дроби, т. е. отношения двух целых чисел. Для древних греков было важно иметь возможность формулировать их геометрические меры (отношения) в терминах (отношений) целых чисел, чтобы оперировать геометрическими величинами в соответствии с правилам арифметики. В основном, идея Евдокса заключалась в том, чтобы дать метод описания отношений длин (т. е. действительных чисел!) в терминах *целых чисел*. Евдоксу удалось сформулировать в рамках операций над целыми числами такие критерии, которые позволяли решать, является ли одно из отношений длин больше другого или их можно считать в точности равными.

В общих чертах идея Евдокса сводится к следующему: если a, b, c и d — четыре длины, то критерием, позволяющим утверждать, что отношение a/b *больше* отношения c/d, будет существование таких целых чисел M и N, что длина a, сложенная сама с собой N раз, больше длины b, сложенной сама с собой M раз, — тогда как длина d, сложенная сама с собой M раз, больше длины c, прибавленной к самой себе N раз[7]. Соответствующий критерий можно аналогичным образом использовать для установления противоположного неравенства $a/b < c/d$. А искомый критерий *равенства* $a/b = c/d$ просто отвечает случаю, когда ни один из двух критериев ($a/b > c/d$ и $a/b < c/d$) не может быть выполнен!

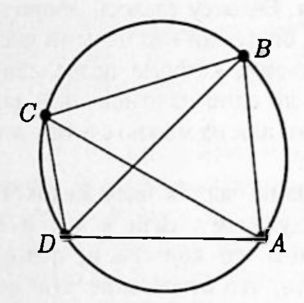
Совершенно точная абстрактная математическая теория действительных чисел

---

[6] Евдокс был также создателем ПОЛЕЗНОЙ теории движения планет, просуществовавшей 2000 лет, развитой позднее Гиппархом и Птолемеем и потому впоследствии получившей название птолемеевой системы!

[7] В современных обозначениях это утверждение означает, что существует дробь, а именно M/N, такая, что $a/b > M/N > c/d$. Такая дробь, лежащая между действительными числами a/b и c/d при условии, что $a/b > c/d$, может быть найдена всегда, поэтому критерий Евдокса действительно выполняется.

была построена только в XIX веке такими математиками, как Дедекинд и Вейерштрасс. Но в действительности, предложенная ими процедура опиралась на те же идеи, которые были открыты Евдоксом примерно двадцатью двумя столетиями раньше! Сейчас нам не обязательно заниматься подробным изучением этой современной теории. Я кратко коснулся ее основных моментов в главе 3 (на с. 98), где для большей наглядности изложения предпочел использовать более привычное десятичное разложение действительных чисел. (В действительности, десятичное разложение была введено Стевином в 1585 году.) Следует также заметить, что хорошо знакомая нам десятичная запись была неизвестна древним грекам.



$AB \cdot CD + AD \cdot BC = AC \cdot BD$

Рис. 5.3. Теорема Птолемея

Однако, между теориями, предложенными Евдоксом с одной стороны, и Дедекиндом и Вейерштрассом — с другой, существует важное различие. Древние греки рассматривали действительные числа как *изначально данные* — в терминах (отношений) геометрических величин — т. е. как свойства «реального» пространства. Древним грекам было необходимо иметь возможность описывать геометрические величины арифметически, чтобы затем в рамках законов и правил арифметики проводить строгие рассуждения над этими геометрическими величинам, а также их суммами и произведениями — существенными составляющими столь многих замечательных геометрических теорем древних. (На рис. 5.3 в качестве иллюстрации приведена знаменитая *теорема Птолемея*, хотя Птолемей открыл ее гораздо позже эпохи, в которую жил Евдокс. Теорема Птолемея устанавливает соотношение, которому удовлетворяют расстояния между четырьмя точками на окружности; в ее формулировке с необходимостью используются как понятие суммы, так и понятие произведения.) Критерии Евдокса оказались необычайно плодотворными и, в частности, позволили древним грекам строго вычислять площади и объемы.

Но для математиков XIX века — и, разумеется, для современных математиков — роль геометрии изменилась. Для древних греков и, в частности, для Евдокса, «действительные» числа были объектами, *извлеченными* из геометрии физического пространства. Ныне мы предпочитаем считать, что действительные числа логически более первичны, чем геометрия. Это позволяет нам конструировать всевозможные *различные* типы геометрии, каждый из которых *исходит* из понятия числа. (Ключевой идеей была идея координатной геометрии, введенная в XVII веке Ферма и Декартом. Координаты можно использовать для *определения* других типов геометрии.) Любая такая «геометрия» должна быть логически непротиворечивой, но не обязательно должна иметь прямое отношение к физическому пространству нашего эмпирического опыта. Конкретную физическую геометрию мы, по-видимому, постигаем через *идеализацию* эмпирического опыта (т. е. в зависимости от наших экстраполяций на бесконечно большие или бесконечно малые размеры, — см. главу 3, с. 102). Проводимые ныне эксперименты достаточно точны и приводят нас с необходимостью к заключению, что наша «извлеченная из эмпирического опыта» геометрия в действительности отличается от евклидова идеала (см. с. 195) и согласуется с геометрией, требуемой в общей теорией относительности Эйнштейна. Однако, несмотря на изменения в наших взглядах на геометрию физического мира, возникших в настоящее время, понятие действительного числа, выдвинутое Евдоксом двадцать три столетия назад, по существу осталось неизменным и является существенным ингредиентом как теории Эйнштейна, так и теории Евклида. В действительности это понятие служит существенным ингредиентом всех современных серьезных физических теорий!

Пятая книга *Начал* Евклида была, по существу, изложением описанной выше «тео-

рии пропорций», введенной Евдоксом. Эта книга имела принципиально важное значение для всего многотомного сочинения Евклида в целом. На самом деле, *Начала* Евклида, впервые увидевшие свет около 300 года до н. э., должны считаться одним из сочинений, оказавших наибольшее влияние в истории человечества. Именно *Начала* Евклида установили эталон для почти всего последующего естественно-научного и математического мышления. Методы *Начал* были дедуктивными, изложение начиналось с четко сформулированных аксиом, которые предполагались «самоочевидными» свойствами пространства; из аксиом выводились многочисленные следствия, многие из которых были важными и поразительными, и совсем не самоочевидными. Не подлежит сомнению, что *Начала* Евклида имели огромное значение для последующего развития естественно-научного мышления.

Величайшим математиком древности несомненно был Архимед (287–212 гг. до н. э.). Остроумно используя теорию пропорций Евдокса, Архимед вычислил площади и объемы многих фигур и тел различной формы, например, сферы и более сложных геометрических форм, в том числе парабол или спиралей. Ныне для этих целей мы использовали бы дифференциальное и интегральное исчисление, но Архимед жил и творил примерно за 19 веков до создания математического анализа, разработанного Ньютоном и Лейбницем! (Можно было бы сказать, что добрая половина — «интегральная» половина — математического анализ была известна еще Архимеду!) Степень математической строгости, достигнутой Архимедом в своих рассуждениях, была безупречной даже по современным стандартам. Работы Архимеда оказали глубокое влияние на математиков и естествоиспытателей последующих веков, в частности, в значительной мере на Галилея и Ньютона. Архимед также ввел (ПРЕВОСХОДНУЮ?) физическую теорию статики (т. е. теорию, занимающуюся изучением законов поведения тел, находящихся в состоянии равновесия, например, законов рычага и законов плавающих тел) и развил статику как дедуктивную науку, аналогично тому, как Евклид изложил науку о геометрическом пространстве и геометрию твердых тел.

Современником Архимеда, которого я также считаю необходимым отметить, был Аполлоний (ок. 262–200 гг. до н. э.), великий геометр, отличавшийся глубиной озарений и остроумием. Ему мы обязаны исследованием теории конических сечений (т. е. эллипсов, парабол и гипербол), которая оказала весьма сильное влияние на Кеплера и Ньютона. Оказалось, что именно эти кривые, что весьма примечательно, необходимы для описания планетных орбит!

Динамика Галилея и Ньютона

Глубоким прорывом, принесенным в естествознание XVII веком, стало понимание *движения*. Древние греки достигли замечательного понимания статики вещей — твердых геометрических тел или тел, находящихся в состоянии *равновесия* (т. е. в состоянии, в котором все действующие на тело силы уравновешены, и движения нет), но не имели хорошего представления о законах, управляющих поведением *реально движущихся* тел. Чего недоставало древним грекам, это хорошей теории *динамики*, т. е. теории, описывающей тот красивый способ, каким природа управляет изменениями положения тел от одного момента времени к другому. Частично (но отнюдь не полностью) это объясняется тем, что у древних греков не было никаких сколь-нибудь точных средств измерения времени, т. е. достаточно хороших «часов». Такие часы необходимы для точного хронометрирования изменений в положении тел. Это позволило бы точно определить скорости и ускорения тел. Наблюдения, произведенные Галилеем в 1583 году, показали, что в качестве надежного средства хранения точного времени можно было бы использовать маятник. Этот факт имел далеко идущие последствия для самого Галилея (и для развития всего естествознания в целом!), так как позволял осуществить точное [4] хронометрирование движения. Примерно через сорок пять лет — с публикацией в 1638 году *Бесед и математических доказательств, касающихся двух новых отраслей науки* Галилея —

начал развиваться новый предмет — *динамика*, и началась трансформация от древнего мистицизма к современной науке!

Позвольте мне выделить всего лишь *четыре* наиболее важные физические идеи, введенные Галилеем. Первая идея Галилея заключалась в том, что сила, действующая на тела, определяет *ускорение*, а не скорость. Что в действительности означают термины «ускорение» и «скорость»? *Скорость* частицы — или какой-нибудь точки тела — это темп изменения во времени положения этой частицы или точки. Скорость обычно принято считать *векторной* величиной, иначе говоря, необходимо принимать во внимание не только величину, но и *направление* скорости (в противном случае мы используем термин «величина скорости», см. рис. 5.4). Ускорение (также векторная величина) — это темп изменения скорости во времени. Таким образом, ускорение в действительности есть *скорость изменения скорости изменения* положения во времени! (Древним было трудно понять сущность понятия «ускорение», так как у них не было адекватных «часов», и они не располагали соответствующими математическими идеями относительно «темпа изменения».) Галилей установил, что сила, приложенная к телу (в случае, исследуемом Галилеем — сила тяжести), управляет ускорением этого тела, но *не* управляет непосредственно его скоростью, как полагали древние, например, Аристотель.



Рис. 5.4. Скорость, величина скорости и ускорение

В частности, в отсутствие приложенной к телу силы его скорость постоянна. Следовательно, неизменяемое движение тела по прямой есть результат *отсутствия* силы (первый закон движения Ньютона). Тела в свободном движении продолжают сохранять состояние равномерного прямолинейного движения, и для того, чтобы они пребывали в этом состоянии, никакой силы не требуется. Действительно, одно из следствий из выведенных Галилеем и Ньютоном законов движения состояло в том, что равномерное прямолинейное движение физически полностью неотличимо от состояния покоя (т. е. отсутствия движения): не существует локального способа, позволяющего отличить равномерное прямолинейное движение от покоя! Галилей особенно четко сформулировал это утверждение (даже более четко, чем Ньютон) и дал ему весьма наглядное описание, использовав образ корабля в море (см. Дрэйк [1953], с. 186–187):

> «Закройтесь вместе с вашим приятелем в кают-компании под палубой большого судна, прихватив с собой мух, бабочек и каких-нибудь других мелких летающих существ. Возьмите также с собой большой сосуд с водой, в котором бы плавала рыбка; подвесьте бутылку, из которой вода капля за каплей вытекала бы в подставленный снизу широкий сосуд. Пока судно будет стоять, внимательно присмотритесь к тому, как мелкие твари летают в каюте с одинаковой быстротой по всем направлениям. Рыбка также плавает одинаково охотно по всем направлениям; капли из бутылки падают в подставленный снизу сосуд... Внимательно пронаблюдав все эти явления, вы пускаетесь в плавание. Судно идет с любой скоростью, какая вам будет угодна. До тех пор и поскольку движение судна будет прямолинейным и равномерным без рысканья то в одну, то в другую сторону, вы не обнаружите ни малейших изменений в наблюденных ранее явлениях и не сможете отличить ни по одному из них, движется ли судно или стоит на месте... Капли будут, как и прежде, падать в подставленный снизу сосуд, ничуть не отклоняясь к корме, хотя пока капли находятся в воздухе, судно успевает пройти значительное расстояние. Рыбка в воде будет плавать вперед (по ходу движения судна) так же часто, как и назад, и с одинаковой легкостью подплывать к корму, в каком бы месте у стенок сосуда он бы ни был насыпан. Наконец,

мухи и бабочки будут по-прежнему летать по всем направлениям, не отдавая предпочтения ни одному из них, не скапливаясь ближе к корме, как бы от усталости, будучи вынужденными следовать курсу судна, от которого они будут отделены на протяжении продолжительных интервалов времени, в течение которых они находятся в воздухе».

Этот замечательный факт, получивший название *принципа относительности Галилея*, имеет в действительности решающее значение для наполнения *копернианской* точки зрения динамическим смыслом. Николай Коперник (1473–1543) и древнегреческий астроном Аристарх (ок. 310–230 гг. до н. э.; не путать с Аристотелем!) за восемнадцать веков до Коперника выдвинули гипотезу о том, что Солнце покоится, а Земля движется, вращаясь вокруг своей собственной оси и обращаясь по орбите вокруг Солнца. Почему мы не ощущаем этого движения, которое происходит со скоростью около нескольких сотен тысяч километров в час? До того, как Галилей выдвинул свою динамическую теорию, этот вопрос действительно представлял настоящую и глубокую загадку для сторонников копернианской картины мироздания. Если бы была верна более ранняя «аристотелевская» версия динамики, согласно которой реальная *скорость* системы в ее движении сквозь пространство влияла бы на динамическое поведение системы, то движение Земли заведомо было бы чем-то непосредственно очевидным для нас. Относительность Галилея позволяет понять, каким образом Земля может находиться в движении, хотя это движение не будет чем-то воспринимаемым нами непосредственно[8].

Заметим, что в рамках галилеевой относительности не существует локального физического смысла, который можно было бы придать понятию «в покое». Это приводит к важным следствиям относительно того, как надлежит рассматривать пространство и время. Интуитивная картина пространства и времени состоит в том, что «пространство» представляет собой своего рода арену, на которой происходят физические события. Физический объект может в один момент времени находиться в одной точке пространства, а в более поздний момент времени может оставаться в той же точке или оказаться в другой точке пространства. Представим себе мысленно, что точки пространства каким-то образом могут сохранять свое положение от одного момента времени до следующего момента так, что имеет смысл говорить о том, изменил ли некоторый объект свое положение в пространстве или не изменил. Но галилеева относительность учит нас, что «состояние покоя» не имеет абсолютного характера и поэтому невозможно придать смысл выражению «одна и та же точка пространства в два различных момента времени». Какая точка евклидова трехмерного пространства физической реальности в один момент времени является «той же» точкой евклидова трехмерного пространства в другой момент времени? На этот вопрос невозможно ответить. Создается впечатление, что для каждого момента времени нам необходимо иметь совершенно «новое» евклидово пространство! Этому можно придать смысл, если рассмотреть *четырехмерную пространственно-временную* картину

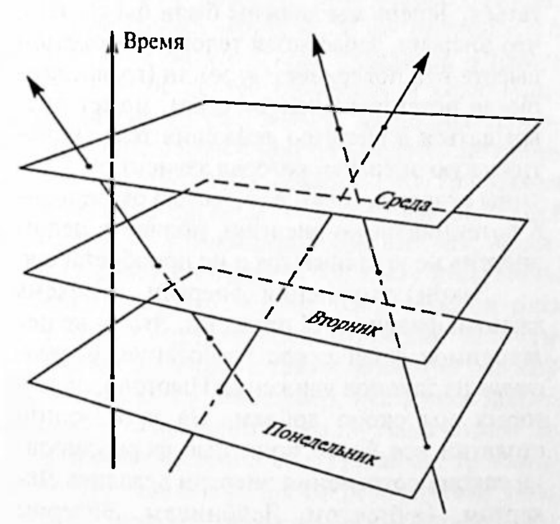


Рис. 5.5. Галилеево пространство-время: частицы, движущиеся равномерно и прямолинейно, изображены в виде прямых

---

[8] Строго говоря, сказанное относится к движению Земли лишь постольку, поскольку его можно считать приближенно *равномерным* и, в частности, без вращения. Действительно, вращательное движение Земли создает (относительно малые) динамические эффекты, которые могут быть обнаружены. Самые заметные из такого рода эффектов — отклонение ветров в северном и южном полушариях в различные стороны. Галилей полагал, что такая неравномерность «ответственна» за приливы.

физической реальности (рис. 5.5). Трехмерные евклидовы пространства, соответствующие различным моментам времени, в этой картине действительно рассматриваются отдельно друг от друга, но все эти пространства объединены, образуя совместно полную картину четырехмерного пространства-времени. Истории частиц, движущихся равномерно и прямолинейно, описываются прямыми (называемыми мировыми линиями) в пространстве-времени. В дальнейшем я еще вернусь к проблеме пространства-времени и относительности движения в контексте эйнштейновской специальной теории относительности. Мы увидим, что довод в пользу четырехмерности обретает в этом случае гораздо бóльшую силу.

Третья из великих догадок Галилея стала ключом к началу понимания *закона сохранения энергии*. Галилея главным образом интересовало движение объектов под действием силы тяжести. Он заметил, что если тело стартует из состояния покоя, то идет ли речь о свободно падающем теле, или о колеблющемся маятнике произвольной длины, или о теле, соскальзывающем по наклонной плоскости, скорость движения всегда зависит только от расстояния по вертикали, пройденного телом от начального положения. Кроме того, достигнутая скорость всегда в точности достаточна для возвращения тела на ту высоту, с которой оно начало двигаться. Теперь мы должны были бы сказать, что энергия, запасенная телом на исходной высоте над поверхностью земли (гравитационная потенциальная энергия), может превращаться в энергию движения тела (кинетическую энергию, которая зависит от величины *скорости* тела), а та, в свою очередь, — в потенциальную энергию, причем в целом энергия не утрачивается и не приобретается.

Закон сохранения энергии — очень важный физический принцип. Это — не независимое физическое требование, а *следствие* из законов движения Ньютона, до которых мы скоро дойдем. На протяжении столетий все более понятные формулировки закона сохранения энергии делались Декартом, Гюйгенсом, Лейбницем, Эйлером и Кельвином. Позднее в этой главе и в главе 7 мы еще вернемся к закону сохранения энергии. Оказывается, что в сочетании с галилеевским принципом относительности закон сохранения энергии приводит к другим законам сохранения, имеющим немалое значение: закону *сохранения массы* и закону *сохранения количества движения* (*импульса*). Количество движения частицы равно произведению ее массы и ее скорости. Знакомые примеры сохранения количества движения возникают при рассмотрении реактивного движения, когда увеличение направленного вперед количества движения ракеты в точности уравновешивается направленным назад количеством движения выхлопных газов (обладающих меньшей массой, но зато большей скоростью). Отдача ружья при выстреле — еще одно проявление закона сохранения количества движения. Еще одним следствием из законов движения Ньютона служит закон *сохранения углового момента* (*момента количества движения*), описывающий постоянство вращения системы вокруг собственной оси. Вращение Земли вокруг собственной оси, равно как и вращение теннисного мяча вокруг собственной оси, не затухают благодаря закону сохранения их угловых моментов. Каждая частица, образующая любое тело, вносит свой вклад в полный угловой момент тела, причем величина этого вклада равна произведению количества движения частицы на расстояние ее от оси вращения (длину перпендикуляра, опущенного из точки, где находится частица, на ось вращения). (Следовательно, угловую скорость свободно вращающегося объекта можно увеличить, сделав объект более компактным. Это приводит к поразительному, но хорошо знакомому действию, часто исполняемому спортсменами на льду и воздушными гимнастами на трапеции. Прижав к себе руки или поджав ноги, они резко увеличивают скорость вращения просто вследствие закона сохранения углового момента!) Как будет показано в дальнейшем, масса, энергия, количество движения (импульс) и угловой момент принадлежат к числу важных для нас понятий. Наконец, мне следовало бы напомнить читателю о пророческой догадке Галилея, понявшего, что в отсутствие атмосферного сопротивления все тела под действием силы тяжести падают с одной и той же скоростью. (Возможно, читатель вспомнит

известную легенду о том, как Галилей сбрасывал с наклонной башни в Пизе по несколько предметов одновременно.) Три столетия спустя то же самое озарение привело Эйнштейна к обобщению принципа относительности на ускоренные системы отсчета и стало, как мы увидим в конце этой главы, краеугольным камнем его необычайной общей теории относительности.

На мощном фундаменте, заложенном Галилеем, Ньютону удалось возвести величественнейший храм. Он сформулировал три закона, управляющие поведением материальных тел. Первый и второй законы Ньютона по существу совпадали с законами, открытыми Галилеем: если на тело не действует никакая сила, то тело продолжает равномерно двигаться по прямой; если на тело действует какая-нибудь сила, то произведение массы тела на ускорение (т. е. скорость изменения количества движения тела) равно этой силе. Заслуга собственно Ньютона состояла в осознании необходимости *третьего* закона движения: сила, с которой тело A действует на тело B, в точности равна по величине и противоположна по направлению силе, с которой тело B действует на тело A (иными словами, «для каждого действия всегда существует равное по величине противодействие»). Три закона движения Ньютона образуют основу основ. «Ньютоновская вселенная» состоит из частиц, движущихся в пространстве, где действуют законы евклидовой геометрии.

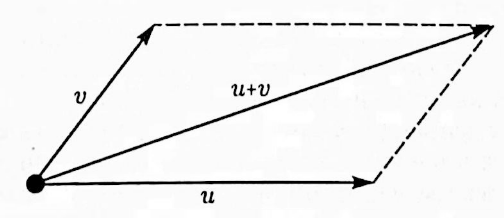


Рис. 5.6. Сложение векторов по правилу параллелограмма

Ускорения этих частиц определяются действующими на них силами. Сила, приложенная к каждой из частиц, получается путем сложения (по правилу *сложения векторов*, см. рис. 5.6) всех сил, действующих на данную частицу со стороны всех *остальных* частиц. Чтобы система была хорошо

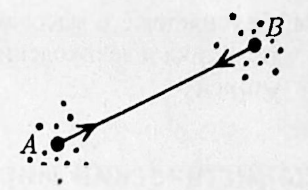


Рис. 5.7. Сила, действующая между двумя частицами, направлена по прямой между ними (и по третьему закону Ньютона сила, действующая на частицу A со стороны частицы B, всегда равна по величине и противоположна по направлению силе, действующей на B со стороны A)

определенной, необходимо задать некоторое четкое правило, которое позволяло бы установить, какая сила действует на частицу A со стороны другой частицы B. Обычно мы требуем, чтобы эта сила действовала по прямой, соединяющей частицы A и B (рис. 5.7). Если речь идет о гравитационной силе, то между A и B возникает сила притяжения, величина которой пропорциональна произведению масс частиц A и B и обратно пропорциональна квадрату расстояния между частицами: *закон обратных квадратов*. Для других типов сил зависимость от взаимного расположения частиц может быть другой, и величина силы в этом случае будет зависеть не от масс частиц, а от какого-то иного их свойства.

Великий Иоганн Кеплер (1571–1630), современник Галилея, заметил, что орбиты планет, описываемые ими вокруг Солнца, имеют форму *эллипсов*, а не окружностей (причем Солнце всегда находится в фокусе, а не в центре эллипса), и сформулировал два других закона, задающих скорости, с которыми планеты движутся по орбитам. Ньютон сумел показать, что три закона Кеплера следуют из его собственной общей модели (с учетом силы притяжения, обратно пропорциональной квадрату расстояния между телами). Кроме того, Ньютон внес многие поправки к кеплеровским эллиптическим орбитам, а также объяснил ряд других эффектов (например, медленное движение оси вращения Земли, замеченное задолго до Ньютона еще древними греками). Чтобы прийти к таким результатам, Ньютону, помимо дифференциального исчисления, пришлось разработать немало дополнительных математических методов. Феноменальный успех, увенчавший эти усилия,

во многом объясняется его высочайшим искусством математика и великолепной физической интуицией.

Механистический мир динамики Ньютона

С введением определенного закона для силы (как обратного квадрата расстояния между телами) ньютоновская модель превращается в точную и определенную систему динамических уравнений. Если положения, скорости и массы различных частиц заданы в некоторый момент времени, то их положения и скорости (равно как и массы, которые считаются *постоянными*) автоматически определены для всех последующих моментов времени. Эта форма *детерминизма*, которой удовлетворяет мир механики Ньютона, оказала (и все еще продолжает оказывать) глубокое влияние на философскую мысль. Попробуем изучить природу ньютонианского детерминизма чуть более подробно. Что он может сказать нам о «свободе воли»? Мог бы в строго ньютонианском мире существовать разум? Найдется ли в нем место хотя бы компьютерам?

Давайте попытаемся представить более конкретно «ньютонианскую» модель мира. Например, мы можем предположить, что частицы материи допустимо считать математическими точками, т. е. объектами, не имеющими никакой пространственной протяженности. В качестве альтернативы все частицы можно считать твердыми сферическими шариками. И в том, и в другом случае нам придется предположить, что законы действия сил, как в случае ньютоновского закона всемирного тяготения, известны. Мы хотим промоделировать и другие встречающиеся в природе силы, такие как *электрические* и *магнитные* взаимодействия (впервые подробно исследованные в 1600 году Уильямом Гильбертом), или сильные *ядерные* взаимодействия, которые, как ныне известно, связывают частицы (протоны и нейтроны), образующие атомные ядра. Электрическое взаимодействие похоже на гравитационное, поскольку тоже удовлетворяет закону обратных квадратов, но при этом одинаково заряженные частицы *отталкивают* (а не притягивают, как в случае гравитационного взаимодействия) друг друга, и величину электрического взаимодействия определяют не массы, а *электрические заряды* частиц. Магнитное взаимодействие, так же как и электрическое, «обратно пропорционально квадрату расстояния»[9], но ядерное взаимодействие имеет совершенно другую зависимость от расстояния: оно очень велико на очень малых расстояниях, сравнимых с внутриатомными, и пренебрежимо мало на бо́льших расстояниях.

Предположим, что мы остановили свой выбор на модели твердых сферических шариков, потребовав, чтобы при столкновении частиц шарики просто идеально *упруго* отражались. Иначе говоря, они должны разлетаться после столкновения без какой бы то ни было потери энергии (или полного количества движения (импульса)), как если бы они были идеальными бильярдными шарами. Нам необходимо также точно задать, какие *силы* должны действовать между шариками. Для простоты мы можем положить, что сила, с которой один шарик действует на любой другой, направлена по прямой, соединяющей центры шариков, а величина силы определяется длиной отрезка между центрами шариков. (Для ньютоновской *гравитации* это предположение выполняется автоматически в силу замечательной теоремы, доказанной Ньютоном; а для других видов сил оно может быть наложено в качестве дополнительного требования.) Если шарики сталкиваются только попарно, а тройные столкновения, как и столкновения более высокого порядка, не происходят, то все вполне определено, и исход столкновения непрерывно зависит от начального состояния (т. е. достаточно малые изменения в начальном состоянии приводят лишь к малым изменениям в конечном). Скользящие столкновения рассматриваются как предельный случай прохождения шариков в непосредственной близости друг от друга. Проблема возникает при рассмотрении

---

[9] Различие между электрическим и магнитным взаимодействиями состоит в том, что индивидуальные «магнитные заряды» (т. е. северные и южные полюсы), по-видимому, не существуют в природе отдельно друг от друга. Магнитные частицы образуют так называемые «диполи», т. е. крохотные магнитики (в которых северный и южный полюсы как бы сливаются вместе).

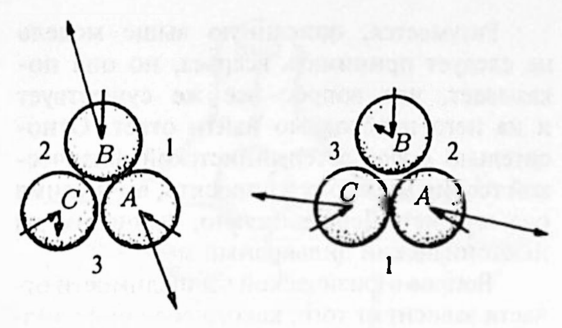


Рис. 5.8. Тройное соударение. Поведение частиц в результате столкновения существенно зависит от того, какие частицы сталкиваются первыми, поэтому исход столкновения не зависит непрерывным образом от начальных данных

тройных столкновений и столкновений более высоких порядков. Например, если происходит одновременное столкновение трех шариков A, B и C, то вся картина значительно меняется в зависимости от того, какое из попарных соударений мы рассматриваем сначала: шарика A с шариком B, а сразу же после этого — C с B; или же мы считаем, что сначала сталкиваются шарики A и C, а затем шарик B сталкивается с шариком A (рис. 5.8). В нашей модели существует *индетерминизм*, когда происходит тройное столкновение! Если угодно, то мы можем просто *исключить* тройные столкновения и столкновения более высокого порядка как «в высшей степени (бесконечно) невероятные». Это дает вполне непротиворечивую схему, но потенциальная проблема тройных столкновений означает, что результирующее поведение частиц может не зависеть непрерывным образом от начального состояния.

Поскольку такое положение дел нас не совсем удовлетворяет, то мы можем отдать предпочтение картине *точечных* частиц. Но для того, чтобы избежать некоторых теоретических трудностей, возникающих в рамках этого подхода (бесконечные силы и бесконечные энергии при столкновении частиц), необходимо сделать дополнительные предположения, в частности о том, что на коротких расстояниях силы, действующие между частицами, всегда становятся отталкивающими. Тогда мы можем обеспечить невозможность столкновения любой пары частиц. (Оно также помогает нам избежать проблемы определения *поведения* частиц при столкновении!) Но для большей наглядности я все-таки буду рассматривать модель твердых сферических шариков, ибо, как мне кажется, подобная «бильярдная» картина для большинства из нас подсознательно как раз и является рабочей моделью *реальности*!

Подчеркнем (игнорируя проблему столкновения нескольких шариков), что ньютонианская [5] бильярдная картина реальности в действительности является *детерминистской* моделью. Слово «детерминистская» надлежит понимать в том смысле, что физическое поведение системы с математической точки зрения полностью определено во все моменты времени в будущем (или в прошлом) положениями и скоростями шариков (во избежание некоторых проблем предположим, что число шариков конечно) в какой-то *один* момент времени. Таким образом, создается впечатление, будто в таком бильярдном мире нет места для разума, который своей «свободной волей» мог бы влиять на поведение материальных объектов. Если мы верим в «свободу воли», то, по-видимому, вынуждены будем усомниться в возможности описания нашего *реального* мира в рамках бильярдной модели.

Мучительный вопрос о «свободе воли» проходит через всю эту книгу — хотя при обсуждении большинства затронутых в ней тем он остается на заднем плане. В этой главе ему предстоит сыграть определенную, но небольшую роль (связанную с проблемой передачи сигналов со сверхсветовой скоростью в теории относительности). Вопросом о свободе воли мы займемся непосредственно в главе 10, и читатель несомненно будет разочарован моим вкладом в эту проблему. Я действительно считаю, что вопрос о свободе воле представляет собой реальную, а не вымышленную проблему — но она в высшей степени нетривиальна и ее трудно сформулировать адекватно. Вопрос о *детерминизме* в физической теории, безусловно, важен, однако я убежден, что он не является камнем преткновения. Например, мир может быть детерминистским, но *невычислимым*. Иначе говоря, будущее может определяться прошлым, но точно рассчитать его при этом будет *в принципе* невозможно. В главе 10

я попытаюсь изложить аргументы, показывающие, что действие нашего наделенного сознанием разума неалгоритмично (т. е. невычислимо). Соответственно, свобода воли, которой мы наделены (по нашему глубокому убеждению), должна быть тесно связана с какой-то невычислимой составляющей законов, управляющих тем миром, в котором мы живем. Независимо от того, принимаем ли мы или отвергаем такую точку зрения на свободу воли, интерес для нас представляет вопрос именно о *вычислимости* данной физической теории (например, ньютоновской динамики), а не о том, является ли она детерминистской. Вопрос о вычислимости отличен от вопроса о детерминизме. Утверждение о том, что это — два совершенно разных вопроса, как раз и служит одним из основных тезисов в данной книге.

Вычислима ли жизнь в бильярдном мире?

Позвольте мне сначала показать на умышленно абсурдном искусственном примере, что вычислимость и детерминизм — понятия различные. Для этого я продемонстрирую «игрушечную модель вселенной», которая детерминистична, но не вычислима. Пусть «состояние» этой вселенной в любой «момент времени» описывается как пара натуральных чисел (m, n). Пусть T_u — фиксированная универсальная машина Тьюринга, например, та, которая описана в главе 2 (с. 80). Чтобы решить, какое состояние этой вселенной наступит в следующий «момент времени», нам необходимо спросить, остановится ли действие машины Тьюринга T_u на m или не остановится (в обозначениях главы 2, с. 82, $T_u(m) \neq \square$ или $T_u(m) = \square$). Если машина Тьюринга T_u останавливается, то состояние в следующий момент времени есть $(m + 1, n)$. Если же машина Тьюринга не останавливается, то состояние в следующий момент времени должно быть $(n+1, m)$. В главе 2 было показано, что не существует алгоритма для решения проблемы остановки машины Тьюринга. Следовательно, не может быть алгоритма предсказания «будущего» в рассматриваемой модели вселенной, несмотря на то, что эта модель вполне детерминистична [6].

Разумеется, описанную выше модель не следует принимать всерьез, но она показывает, что вопрос все же существует и на него необходимо найти ответ. Относительно *любой* детерминистской физической теории мы сможем спросить, вычислима она или нет. Действительно, вычислим ли ньютонианский бильярдный мир?

Вопрос о физической вычислимости отчасти зависит от того, какого рода информацию о данной системе мы хотим получить. Я могу придумать целый ряд вопросов о конкретной физической системе, на которые — как мне *кажется* — в случае ньютоновской бильярдной модели не существует вычислимого (т. е. алгоритмически получаемого) ответа. Одним из таких вопросов мог бы быть следующий: столкнется ли когда-нибудь шарик A с шариком B? Имеется в виду, что в качестве *начальных условий* нам в некоторый момент времени ($t = 0$) задаются положения и скорости всех шариков; и задача состоит в том, чтобы, исходя из этих данных, выяснить, сталкиваются или не сталкиваются шарики A и B в некоторый последующий момент времени ($t > 0$). Чтобы придать задаче бóльшую конкретность (хотя и сделав ее при этом не особенно реалистичной), мы можем предположить, что все шарики имеют одинаковый радиус и одинаковую массу и что, скажем, сила, действующая между шариками, обратно пропорциональна квадрату расстояния между ними. Одна из причин, по которой я сделал предположение о невозможности алгоритмически получить ответ на этот вопрос, заключается в том, что сама модель несколько напоминает «бильярдную модель для вычисления», предложенную Эдвардом Фредкином и Томмазо Тоффоли (Фредкин, Тоффоли [1982]). В их модели шарики (вместо того, чтобы попарно взаимодействовать по закону обратных квадратов) были ограничены различными «стенками», но упруго отражались при столкновениях друг с другом — по аналогии с теми ньютоновскими шариками, которые я только что описывал (рис. 5.9). В модели Фредкина—Тоффоли все основные логические операции компьютера могут выполняться с помощью шариков. Модель позволяет имитировать вычисления, производимые любой машиной Тьюринга: конкретный

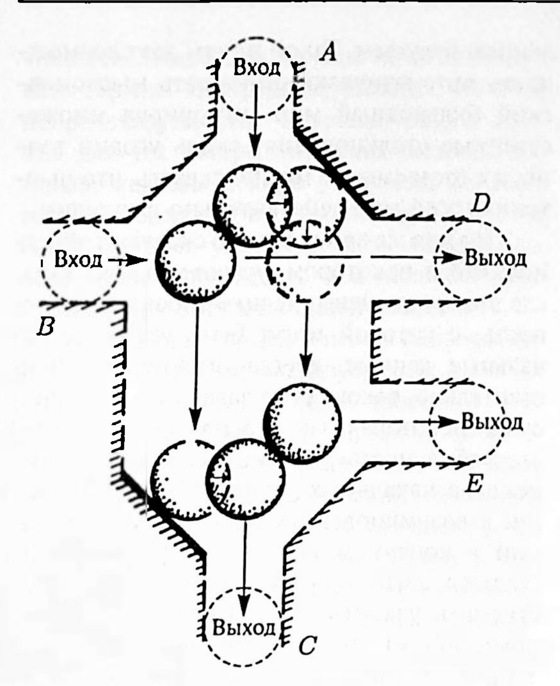


Рис. 5.9. «Переключатель» (конструкции А. Ресслера) в компьютере Фредкина—Тоффоли на бильярдных шарах. Если шар попадает в переключатель через вход B, то в дальнейшем он покидает переключатель через выход D или E в зависимости от того, попадает ли другой шар в переключатель через вход A (предполагается, что шары попадают в переключатель через входы A и B одновременно)

выбор машины Тьюринга T_u определяет конфигурацию «стенок» и т. д. в машине Фредкина—Тоффоли; начальное состояние движущихся шариков соответствует информации на входной ленте машины Тьюринга; а содержимое на выходной ленте соответствует конечному состоянию шариков. Таким образом, можно, в частности, спросить: останавливается ли когда-нибудь такая-то и такая-то машина Тьюринга? «Остановка» может быть сформулирована как состояние при котором шарик A сталкивается, в конце концов, с шариком B. То, что на этот вопрос невозможно ответить алгоритмически (с. 83), по крайней мере *наводит на мысль* о том, что ньютоновский вопрос «сталкивается ли когда-нибудь шарик A с шариком B?», который был поставлен мной первоначально, тоже не может быть разрешен алгоритмически.

В действительности, ньютоновская задача является гораздо более каверзной, чем задача, поставленная Фредкином и Тоффоли. Эти авторы могли задавать состояние своей модели с помощью *дискретных* параметров (т. е. при помощи утверждений «да или нет» типа «шарик либо находится в данном туннеле, либо не находится»). Но в полной ньютоновской задаче начальные положения и скорости шариков необходимо задавать с бесконечной точностью в терминах координат, которые являются *действительными числами*, а не принимают дискретные значения. Таким образом, мы снова сталкиваемся со всеми проблемами, которые нам уже приходилось рассматривать, когда в главе 4 мы пытались ответить на вопрос, рекурсивно ли множество Мандельброта. Что *означает* «вычислимость», когда в качестве входных и выходных данных допускаются непрерывно изменяющиеся параметры? [7] Проблему можно слегка облегчить, предположив, что все начальные положения и скорости заданы *рациональными* числами (хотя нельзя ожидать, что координаты и компоненты скорости останутся рациональными в более поздние рациональные моменты времени t). Напомним, что рациональное число представимо в виде отношения двух целых чисел и, следовательно, определяется в дискретных конечных терминах. Используя рациональные числа, мы можем сколь угодно точно аппроксимировать любые наборы начальных данных, которые собираемся использовать в своих вычислениях. И предположение о том, что при рациональных начальных данных может не существовать алгоритма, позволяющего определить, столкнутся в конце концов или нет шарики A и B, — отнюдь не лишено смысла.

Однако на самом деле, когда говорят: «Ньютонианский бильярдный мир не вычислим», имеют в виду совсем другое. Та модель, которую я сравниваю с ньютонианским бильярдным миром — а именно, «бильярдный компьютер» Фредкина—Тоффоли — действует как вычислительный алгоритм. В конечном счете, это и было квинтэссенцией идеи Фредкина и Тоффоли — что их модель должна вести себя как (универсальный) компьютер! Вопрос, который я пытаюсь сейчас прояснить, сводится к следующему: можно ли представить себе, что человеческий мозг, используя некоторые подходящие «невычислимые» физические

законы, работает в определенном смысле «лучше», чем машина Тьюринга? Бесполезно пытаться использовать что-нибудь вроде следующего утверждения:

> «Если шарик A никогда не сталкивается с шариком B, то ответ на Ваш вопрос будет „нет"».

Чтобы окончательно удостовериться в том, что шарик A действительно никогда не сталкивается с шариком B, пришлось бы прождать вечность! Разумеется, машины Тьюринга *ведут себя* именно так.

На самом деле, существуют, по-видимому, достаточно весомые указания в пользу своего рода *вычислимости* ньютонианского бильярдного мира (по крайней мере, если оставить в стороне проблему множественных столкновений). Способ, которым мы пользуемся для того, чтобы рассчитать поведение такого мира, сводится к введению аппроксимаций. Мы могли бы предположить, что центры шариков по определению располагаются в узлах некоторой точечной решетки, причем координаты узлов измерены, например, с точностью до сотых долей единицы. Время также можно считать «дискретным»: все допустимые моменты времени должны быть кратными некоторой небольшой единице (обозначаемой, скажем, Δt). Это приводит к разным дискретным возможностям для «скоростей» (разностей между значениями положений точек на решетке в два последовательных разрешенных момента времени, деленных на Δt). Соответствующие приближения для ускорений вычисляются с использованием закона силы, и, в свою очередь, используются для получения значений «скоростей». После чего с требуемой точностью вычисляются новые положения шариков в узлах решетки в следующий допустимый момент времени. Вычисления производятся до тех пор, пока сохраняется указанная точность. Вполне может оказаться, что точность будет потеряна раньше, чем мы успеем рассчитать состояние системы для достаточно большого числа моментов времени. В этом случае процедура начинается снова со значительно более мелкой пространственной решеткой и более частыми допустимыми моментами времени. Это позволяет достичь большей точности — и рассчитать поведение системы в более отдаленном будущем. Такой прием дает возможность математически описывать ньютоновский бильярдный мир (игнорируя множественные столкновения) сколь угодно точно, и в этом смысле можно сказать, что ньютонианский мир действительно вычислим.

Но в то же время можно сказать и обратное: что в некотором (*практическом*) смысле этот мир «невычислим», поскольку точность, с которой могут быть *известны* начальные данные, всегда ограничена. Действительно, такого рода задачам всегда присуща некоторая (и весьма значительная) «нестабильность». Очень небольшое изменение в начальных условиях может привести к возникновению чудовищных изменений в конечном состоянии. (Всякий, кто пытался загнать в лузу бильярдный шар, стремясь ударить его промежуточным шаром, поймет, что я имею в виду!) Сказанное становится очевидным, когда происходят (последовательные) столкновения, но такие неустойчивости в поведении могут встречаться и в случае действия ньютоновского тяготения на расстоянии (если гравитирующих тел больше двух). Для обозначения этого типа неустойчивости часто используется термин «хаос», или «хаотическое поведение». Например, хаотическое поведение важно, когда речь заходит о погоде. Хотя ньютоновские уравнения, управляющие стихиями, хорошо изучены, долговременный прогноз погоды печально известен своей ненадежностью!

Все это не похоже на тот тип «невычислимости», который можно было бы каким-то образом «использовать». Невычислимость в данном случае обусловлена просто тем, что из-за существования предела точности, с которой может быть известно начальное состояние, будущее состояние в принципе не поддается точному расчету на основании известных начальных условий. На самом деле, в этом случае к будущему поведению системы примешивается *случайный элемент* — и только. Если же работа мозга все-таки опирается на *полезные* невычислимые составляющие физических законов, то последние должны быть совершенно другими — и более конструктивными — по своей природе. Поэтому я не буду называть «хаотическое» поведение такого рода

«невычислимостью», предпочитая использовать термин «непредсказуемость». Наличие непредсказуемости — весьма общее явление для тех детерминистских законов, которые, как мы вскоре убедимся, действительно возникают в (классической) физике. Но мы скорее уж предпочтем *минимизировать* непредсказуемость, чем «использовать» ее в конструкции мыслящей машины!

Обсуждая в общем и целом вопросы вычислимости и непредсказуемости, нам будет полезно принять более широкую, чем прежде, точку зрения на природу физических законов. Это позволит рассматривать не только схему ньютоновской механики, но и более поздние теории, пришедшие ей на смену. И сперва нам стоит окинуть беглым взглядом замечательную формулировку законов механики, предложенную Гамильтоном.

Гамильтонова механика

Своими успехами ньютоновская механика обязана не только своей способности исключительно точно описывать физический мир, но и обилию порожденных ею математических теорий. Замечательно, что все ПРЕВОСХОДНЫЕ теории природы оказались весьма щедрыми источниками математических идей. В этом кроется глубокая и прекрасная тайна: все наиболее точные теории в то же время необычайно плодотворны и с точки зрения *математики*. Не подлежит сомнению, что это свидетельствует о каких-то глубоких связях между реальным окружающим нас миром и платоновским миром математики. (Далее, (в главе 10, с. 367) я постараюсь еще раз вернуться к этому вопросу.) Возможно, ньютоновская механика в этом отношении не имеет себе равных, так как ее рождение привело к возникновению дифференциального и интегрального исчисления. Кроме того, специфическая ньютонианская схема дала рождение массе замечательных математических идей, составляющих *классическую механику*. Имена многих великих математиков XVIII и XIX веков связаны с развитием этой науки: Эйлер, Лагранж, Лаплас, Лиувилль, Пуассон, Якоби, Остроградский, Гамильтон. То, что принято называть «гамильтоновой теорией» [8], включает в себя многое из проделанной ими работы. Сейчас мы вкратце коснемся общих положений этой теории. Разносторонний и самобытный ирландский математик Уильям Роуан Гамильтон (1805–1865), автор гамильтоновых циклов (обсуждаемых на с. 142), придал этой теории такую форму, которая особо подчеркивала аналогию с распространением волн. Это указание на существование взаимосвязи между волной и частицей (равно как и форма самих уравнений Гамильтона) сыграло важную роль в последующем развитии *квантовой механики*. К этой стороне дела я еще вернусь в следующей главе.

В рамках гамильтоновой теории впервые появились «переменные» для описания физической системы. До Гамильтона *положения* частиц считались первичными, а скорости считались просто быстротой изменения положения частиц во времени. Напомним, что для задания начального состояния ньютоновской системы нам необходимы положения *и* скорости всех частиц — только тогда мы можем определить последующее поведение системы. В рамках гамильтоновой формулировки необходимо выбирать *импульсы*, а не скорости частиц. (На с. 158 мы отметили, что импульс частицы есть не что иное, как произведение ее скорости на массу.) Само по себе это нововведение может показаться несущественным, но важно здесь другое: положение и импульс каждой частицы в гамильтоновой формулировке надлежит рассматривать как *независимые*, более или менее равноправные величины. Тем самым, используя гамильтонову формулировку, мы «делаем вид», что импульсы различных частиц не имеют никакого отношения к быстроте изменения переменных, описывающих их относительное положение, а представляют собой отдельный набор переменных — и, как следствие, мы можем считать импульсы совершенно независимыми от изменения положений движущихся частиц. В гамильтоновой формулировке мы располагаем *двумя* системами уравнений: одна из них говорит нам о том, как изменяются во времени *импульсы* различных частиц, другая — о том,

как изменяются во времени *положения* частиц. И в том, и в другом случае быстрота изменений определяется различными положениями и импульсами в рассматриваемый момент времени.

Грубо говоря, первая система гамильтоновых уравнений выражает второй, самый важный закон движения Ньютона (быстрота изменения импульса = силе), тогда как вторая система уравнений Гамильтона говорит нам о том, чему *равны* импульсы, выраженные в терминах скоростей (быстрота изменения положения = импульс/массу). Напомним, что в формулировках законов движения Галилея—Ньютона использовались ускорения (или быстрота изменения быстроты изменения положения, т. е. уравнения «второго порядка»), тогда как в гамильтоновой формулировке нам достаточно говорить только о быстроте изменения величин (уравнения «первого порядка»). Все гамильтоновы уравнения выводятся всего лишь из одной важной величины: функции Гамильтона H, представляющую собой *полную энергию* системы, выраженную в переменных, описывающих положения и импульсы.

Гамильтонова формулировка дает весьма изящное и симметричное описание механики. Выпишем здесь гамильтоновы уравнения просто для того, чтобы понять, как они выглядят, хотя многие читатели, возможно, и не знакомы с принятыми в математическом анализе обозначениями, необходимыми для полного понимания — впрочем, оно сейчас и не требуется. Все, что нам сейчас действительно нужно знать о дифференциальном исчислении, ограничивается пониманием смысла «точки» в левых частях уравнений Гамильтона — она означает *быстроту изменения по времени* (в первом случае — импульса, во втором случае — положения):

$$\dot{p}_i = -\frac{\partial H}{\partial x_i}, \quad \dot{x}_i = \frac{\partial H}{\partial p_i}.$$

Индекс i здесь использован просто для того, чтобы отличать все различные координаты импульсов $(p_1, p_2, p_3, p_4, \ldots)$ и положений $(x_1, x_2, x_3, x_4, \ldots)$. Для n частиц, не ограниченных наложенными на них связями, мы получаем $3n$ координат импульсов и $3n$ координат положений (по одной координате для каждого из трех независимых направлений в пространстве). Символ ∂ относится к операции «частного дифференцирования» (взятию производной по одной переменной при сохранении постоянных значений всех остальных переменных), а H, как сказано выше, означает функцию Гамильтона. (Если Вы ничего не знаете о «дифференцировании» — не стоит беспокоиться. Просто рассматривайте правые части уравнений Гамильтона как некие вполне определенные математические выражения, записанные через x_i и p_i.)

Координаты $x_1, x_2, \ldots$ и $p_1, p_2, \ldots$ могут на самом деле использоваться для обозначения более общих вещей, а не только обычных декартовых координат для частиц (т. е. когда x_i — обычные расстояния, измеряемые по трем различным направлениям, расположенным под прямыми углами друг к другу). Например, некоторые из x_i в гамильтоновом случае можно считать *углами* — тогда соответствующие p_i превращаются в *угловые* моменты (см. с. 254) вместо импульсов — или вообще какими-нибудь совершенно абстрактными величинами. Замечательно, что при этом гамильтоновы уравнения по-прежнему сохраняют в точности ту же форму. Действительно, при подходящем выборе функции Гамильтона H гамильтоновы уравнения остаются в силе для *любой* системы классических уравнений, а не только для уравнений Ньютона. В частности, они выполняются для теории Максвелла(—Лоренца), к рассмотрению которой мы вскоре приступим. Гамильтоновы уравнения можно записать и для специальной теории относительности. Даже общую теорию относительности (при соблюдении должной осторожности) можно представить в гамильтоновой форме. Кроме того, как мы убедимся в дальнейшем при знакомстве с уравнением Шредингера (с. 254), гамильтонова формулировка служит отправным пунктом для вывода уравнений квантовой механики. Такое единство формы в структуре динамических уравнений, сохранившееся несмотря на все революционные новшества, введенные в физические теории за минувшие столетия, поистине удивительна!

Фазовое пространство

Форма гамильтоновых уравнений позволяет нам «наглядно представить» эволюцию классической системы, используя весьма мощный и универсальный подход. Попытаемся вообразить «пространство» большого числа измерений, по одному измерению на каждую из координат $x_1, x_2, \ldots p_1, p_2, \ldots$. (Математические пространства часто имеют размерность выше трех.) Такое пространство называется *фазовым пространством* (рис. 5.10). Для n свободных частиц размерность фазового пространства равна $6n$ (по три координаты положения и по три координаты импульса для каждой частицы). Читателя может обеспокоить то, что даже для *одной-единственной* частицы размерность фазового пространства оказывается вдвое большей, чем мы обычно привыкли представлять! Но секрет успеха заключается в том, чтобы не пасовать перед трудностями. Конечно, шестимерное пространство действительно имеет бóльшую размерность, чем та, которую можно с ходу (!) представить — но даже если бы могли себе его представить, то пользы от этого оказалось бы немного. Например, всего лишь для комнаты, полной молекул газа, размерность фазового пространства могла бы равняться, например, такой величине:

10 000 000 000 000 000 000 000 000 000.

Попытка наглядно представить себе пространство столь высокой размерности заранее обречена на провал! Так что лучше даже и не пытаться делать это даже в случае фазового пространства для одной-единственной частицы. Просто представьте себе несколько расплывчатую трехмерную (или даже всего лишь двумерную) область. Взгляните еще раз на рис. 5.10. Этого вполне достаточно.

А как теперь наглядно представить себе уравнения Гамильтона для фазового пространства? Прежде всего следует помнить о том, что на самом деле изображает *одна точка* Q фазового пространства. Она соответствует некоторому конкретному набору значений всех координат положений $x_1, x_2, \ldots$ и всех координат импульсов p_1, $p_2, \ldots$. То есть, точка Q представляет *всю* нашу *физическую систему* в определенном состоянии движения, заданного для каждой из образующих ее частиц в отдельности. Уравнения Гамильтона говорят нам о степени быстроты изменения всех этих координат, если их текущие значения известны, т. е. управляют движениями всех отдельных частиц. В переводе на язык фазового пространства уравнения Гамильтона описывают дальнейшее поведение точки Q в этом пространстве, если нам задано ее текущее положение. Таким образом, в каждой точке фазового пространства мы имеем маленькую стрелку (точнее: *вектор*), которая говорит нам о том, как движется точка Q — а это позволяет описывать эволюцию во времени

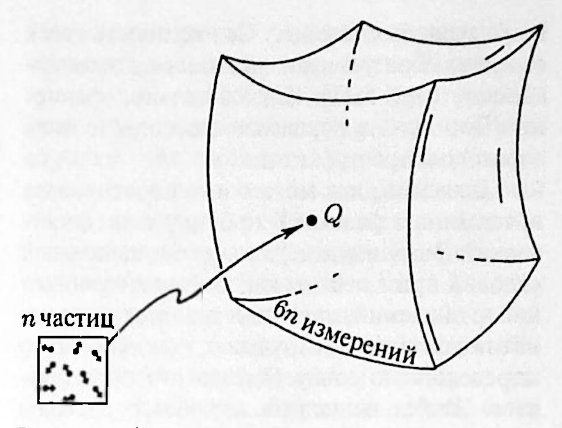


Рис. 5.10. Фазовое пространство. Каждая точка Q фазового пространства описывает полное состояние некоторой физической системы, включающее в себя мгновенные движения всех ее частей

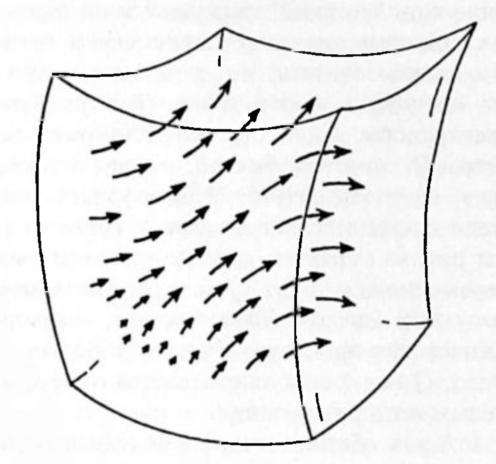


Рис. 5.11. Векторное поле в фазовом пространстве, представляющее эволюцию системы во времени в соответствии с уравнениями Гамильтона

всей нашей системы. Совокупность всех стрелок образует так называемое *векторное поле* (рис. 5.11). Следовательно, уравнения Гамильтона определяют векторное поле в фазовом пространстве.

Выясним, как можно интерпретировать в терминах фазового пространства физический *детерминизм*. В качестве начальных условий при $t = 0$ мы имели бы конкретный набор значений, заданных для всех координат положений и импульсов, т. е. некоторую определенную точку Q фазового пространства. Чтобы вычислить эволюцию системы во времени, надо просто следовать стрелкам. Таким образом, все поведение нашей системы (независимо от степени ее сложности) описывается в фазовом пространстве всего лишь одной точкой, движущейся по стрелкам, которые она встречает на своем пути. Мы можем считать, что стрелки указывают «скорость» нашей точки Q в фазовом пространстве. Если стрелка «длинная», то точка Q движется быстро, а если «короткая» — то медленно. Чтобы узнать, что наша система делает в момент времени t, мы просто смотрим, куда к этому времени переместилась точка Q, следуя указаниям попутных стрелок. Ясно, что это — детерминистская процедура. Характер движения точки Q полностью определяется гамильтоновым векторным полем.

А как обстоит дело с вычислимостью? Если мы стартовали из вычислимой точки фазового пространства (т. е. из точки, у которой все координаты положения и импульсов являются вычислимыми числами, см. главу 3, с. 98), и с момента начала движения прошло вычислимое время t — то закончим ли мы с необходимостью в точке, которая может быть вычислимым образом получена из t и исходных значений координат? Ответ, очевидно, зависит от выбора функции Гамильтона H. Действительно, в функцию H могут входить *физические константы* — такие, как ньютоновская постоянная тяготения или скорость света, величина которых зависит от выбора единиц; или другие, описывающиеся точными числовыми выражениями — и поэтому, чтобы положительно ответить на поставленный вопрос, необходимо сначала убедиться в том, что все эти постоянные *вычислимы*. В *таком* случае я осмелюсь *предположить*, что для обычных гамильтонианов (т. е. функций H), встречающихся в физике, ответ может быть утвердительным. Но это — всего лишь догадка, и вопрос — интересный вопрос! — остается пока открытым. Надеюсь, что со временем он будет изучен более основательно.

С другой стороны, мне кажется, — по тем же самым причинам, которых я кратко коснулся в связи с бильярдным миром — что этот вопрос не настолько существенен. Ведь чтобы утверждение о невычислимости точки фазового пространства имело смысл, необходимо было бы задавать ее координаты с *бесконечной* точностью, т. е. со всеми десятичными знаками после запятой! (Число, записываемое *конечным* количеством десятичных знаков, всегда вычислимо.) Конечный отрезок десятичного разложения любого числа ничего не говорит нам о возможности вычислить оставшуюся часть. Но точность всех физических измерений ограничена возможностями приборов, поэтому они могут дать нам информацию лишь о конечном числе знаков десятичного разложения. Обесценивает ли это само понятие «вычислимого числа» применительно к физическим измерениям?

Действительно, если мы рассматриваем устройство, которое могло бы использовать каким-нибудь *полезным* образом некие (гипотетические) невычислимые составляющие физических законов, то разумно предположить, что оно не должно зависеть от произведения измерений с неограниченной точностью. Но возможно, я сейчас стараюсь рассуждать слишком строго. Предположим, что у нас имеется физическое устройство, которое в силу известных теоретических причин реализует некоторую интересную математическую процедуру неалгоритмического характера. Тогда поведение этого устройства — при условии, что мы имеем возможность точно удостовериться в этом — позволило бы получать правильные ответы на последовательность математически содержательных вопросов, для решения которых не существует алгоритма (подобно вопросам, рассмотренным в главе 4). Любой *наперед заданный* алгоритм на определенной стадии такого процесса дал бы сбой — тогда как наше

устройство *на той же стадии* выдало бы некоторый новый результат. Действительно, это устройство могло бы осуществлять изучение некоторого физического параметра со все большей и большей точностью, необходимой для дальнейшего продвижения по списку вопросов. Однако мы *действительно* получим нечто новое от нашего устройства на какой-то *конечной* стадии точности, по крайней мере пока нам не удастся найти усовершенствованный алгоритм для ответа на указанную последовательность вопросов: затем нам следовало бы повысить точность, чтобы продвинутся еще дальше — до тех пор, пока наш *усовершенствованный* алгоритм не окажется бессилен.

Тем не менее, создается впечатление, что даже все возрастающая точность в определении физического параметра неудобна в качестве способа кодирования информации. Гораздо предпочтительнее было бы получать нашу информацию в «дискретной» (или «цифровой») форме. В этом случае ответы на вопросы, расположенные все дальше и дальше от начала списка, могли бы быть получены путем рассмотрения все большего количества дискретных единиц или, быть может, путем повторного рассмотрения некоторого *фиксированного* набора дискретных единиц, где требуемая неограниченная информация распределялась бы по все более длинным временным интервалам. (Мы могли бы представить себе, что эти дискретные единицы построены из частей, каждая из которых может находиться в одном из двух состояний — «вкл.» или «выкл.» — подобных единицам и нулям в описании машины Тьюринга, приведенном в главе 2.) Для этого нам, как представляется, требуются такие устройства, которые могли бы принимать (отличимые) дискретные состояния и, совершив определенные эволюции в соответствии с динамическими законами, снова перейти в один из наборов дискретных состояний. Если бы это было так, то мы могли бы избежать необходимости изучать каждое устройство с произвольно высокой степенью точности.

Возникает вопрос: действительно ли гамильтоновы системы ведут себя подобным образом? Необходимым условием для этого, видимо, должна быть некоторая устойчивость в поведении системы, позволяющая четко устанавливать, в каком из таких дискретных состояний находится наше устройство. При этом желательно будет зафиксировать это состояние (по крайней мере на некоторый достаточно продолжительный период времени) и добиться того, чтобы оно (устройство) не дрейфовало из одного состояния в другое. Кроме того, если система оказывается в этих состояниях с небольшой погрешностью, то нам бы не хотелось, чтобы погрешности накапливались; наоборот: мы будем требовать, чтобы такие погрешности со временем *сглаживались*. К тому же, наше искомое устройство должно было бы состоять из частиц (или каких-то других подэлементов), которые с необходимостью описывались бы в терминах непрерывных параметров, причем каждое отличимое «дискретное» состояние покрывало бы некоторый *диапазон* значений этих непрерывных параметров. (Например, можно представлять разные дискретные состояния с помощью частицы, лежащей либо в одном, либо в другом ящике. Чтобы указать, что частица действительно находится в одном из них, мы будем говорить, что координаты положения частицы лринадлежат определенному диапазону значений.) С точки зрения фазового пространства это означает, что каждая из «дискретных» альтернатив должна соответствовать некоторой *области* в фазовом пространстве так, чтобы различные точки фазового пространства, принадлежащие одной и той же области, отвечали бы *одному и тому же* состоянию нашего устройства (рис. 5.12).

Предположим теперь, что наше устройство стартует из точки фазового пространства, принадлежащей некоторой области R_0, которая соответствует одной из таких возможностей. Мы будем считать, что область R_0 перемещается вдоль гамильтонова векторного поля до тех пор, пока в момент времени t она не переходит в область R_t. Представляя себе такое развитие событий, мы тем самым описываем эволюцию нашей системы во времени при *всех* возможных начальных состояниях, соответствующих одной и той же альтернативе (рис. 5.13). Вопрос об *устойчивости* (в том смысле, в каком мы трактуем устойчивость здесь) сводится

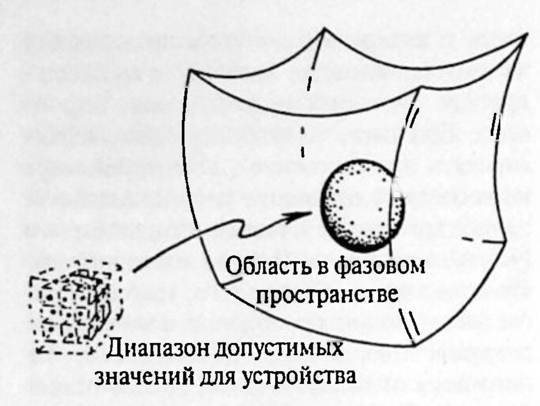


Рис. 5.12. *Область* в фазовом пространстве соответствует диапазону возможных значений пространственных координат и импульсов всех частиц. Такая область может представлять отдельное отличимое состояние (т. е. «альтернативу») какого-нибудь устройства

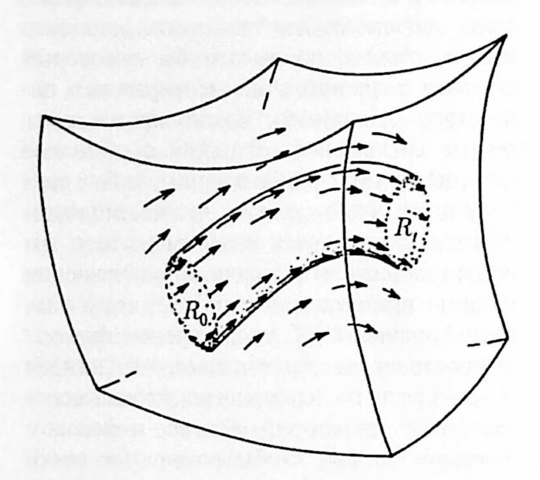


Рис. 5.13. С течением времени область R_0 фазового пространства, увлекаемая вдоль векторного поля, переходит в новую область R_t. Это может служить описанием эволюции во времени некоторого определенного состояния нашего устройства

к вопросу о том, остается ли с ростом t область R_t локализованной или начинает расплываться по всему фазовому пространству. Если область R_t со временем сохраняет конечный объем, то мы будем говорить, что наша система демонстрирует устойчивое поведение. Точки фазового пространства, близкие друг к другу (настолько, что они соответствуют конкретным физическим состояниям системы, которые существенно похожи друг на друга), остаются близкими, и погрешности в указании их положения со временем не увеличиваются. Любое чрезмерно сильное расплывание начальной области R_0 в результате приводит к появлению непредсказуемой составляющей в поведении системы.

А что вообще можно сказать о гамильтоновых системах? Стремятся ли области фазового пространства расплываться со временем или все-таки нет? Казалось бы, при такой общей постановке проблемы сказать о ней можно будет немного. Однако для гамильтоновых систем существует весьма красивая теорема, принадлежащая выдающемуся французскому математику Жозефу Лиувиллю (1809–1882), которая утверждает, что *объем* любой области фазового пространства должен оставаться постоянным при любых изменениях состояния системы, происходящих в соответствии с уравнениями Гамильтона. (Разумеется, размерность «объема» следует понимать в смысле размерности фазового пространства.) Следовательно, объем каждой области R_t должен быть таким же, как объем исходной области R_0. На первый взгляд теорема Лиувилля позволяет утвердительно ответить на вопрос об устойчивости гамильтоновых систем. В силу того, что *размер* исходной области (в смысле ее объема в фазовом пространстве) *не может возрастать*, создается впечатление, будто наша исходная область не может со временем расплываться по всему фазовому пространству.

Однако такое впечатление обманчиво, и, немного поразмыслив над этим, мы поймем, что в действительности может произойти прямо противоположная ситуация! На рис. 5.14 я попытался наглядно изобразить такое поведение системы, которое можно было бы ожидать в общем случае. Представим себе, что начальная область R_0 невелика и имеет «приемлемую» форму — достаточно гладкую, лишенную причудливых выступов — которая указывает на то, что при описании состояний, принадлежащих этой области, чрезмерно высокая точность совсем необязательна. Но с течением времени область R_t начинает деформироваться и растягиваться — сначала принимая форму, напоминающую амебу, а затем образуя причудливые отростки, которые простираются далеко в стороны, замысловато извиваясь то в одном, то в другом направлении.

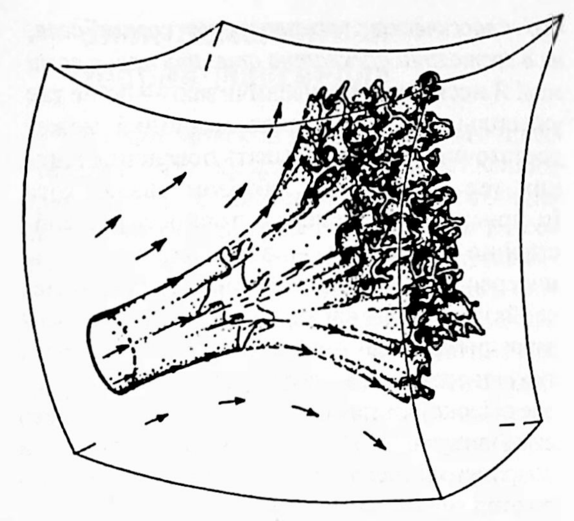


Рис. 5.14. Несмотря на то, что — согласно теореме Лиувилля — объем фазового пространства сохраняется постоянным, он, как правило, будет расплываться в результате чрезвычайно сложной эволюции системы во времени

Объем при этом действительно сохраняется, но тот же самый объем может теперь истончиться и распределиться по обширной области фазового пространства. Практически аналогичная картина будет наблюдаться в случае с капелькой чернил, попавшей в большую емкость с водой. В то время, как реальный объем чернильной жидкости остается неизменным, она постепенно истончается, распределяясь по всему объему емкости. Вероятно, подобным образом ведет себя и исходная область R_0 в фазовом пространстве. Она не обязательно должна расплываться по *всему* фазовому пространству (эта предельная ситуация известна под названием «эргодической») — но вполне может в конце концов занять область, значительно превышающую ее первоначальный объем. (Дальнейшее обсуждение см. в книге: Дэвис [1974].)

Трудность заключается в том, что сохранение объема отнюдь не влечет за собой сохранение *формы*: малые области имеют тенденцию деформироваться, и их деформации простираются на большие расстояния. В многомерных пространствах проблема расплывания начальной области гораздо более серьезна, чем в пространствах малой размерности, так как «направлений», по которым расплываются отдельные части нашей области, гораздо больше. На самом деле, вместо того, чтобы «помочь» нам держать область R_t под контролем, теорема Лиувилля создает фундаментальную проблему! Не будь теоремы Лиувилля, можно было бы представить, что бесспорная тенденция к расплыванию области в фазовом пространстве могла бы (при соответствующих обстоятельствах) компенсироваться уменьшением полного объема. Но теорема Лиувилля говорит нам, что такое уменьшение *невозможно*, и нам остается только мириться с таким поразительным свойством — универсальным для всех классических динамических (гамильтоновых) систем нормального типа! [9]

Помня о неизбежном расплывании исходной области в фазовом пространстве, уместно спросить: а как в таком случае вообще возможно делать предсказания в классической механике? Это действительно непростой вопрос. Расплывание начальной области говорит нам о том, что независимо от степени точности, с которой мы знаем начальное состояние системы (конечно, в разумных пределах), тенденция к возрастанию погрешностей со временем сделает нашу исходную информацию практически бесполезной. В этом смысле классическая механика в принципе *непредсказуема*. (Вспомним введенное выше понятие «хаоса».)

Чем же в таком случае объяснить явный успех ньютоновской механики? Говоря о небесной механике (т. е. движении небесных тел под действием сил гравитации), в качестве наиболее вероятной причины можно назвать, наверное, то, что, во-первых, небесная механика занимается изучением сравнительно небольшого числа связанных тел (Солнца, планет и их естественных спутников — лун), между которыми имеется большой разброс по массе, поэтому в первом приближении возмущающим действием менее массивных тел на более массивные можно пренебречь и рассматривать только взаимодействие *нескольких* массивных тел друг на друга; во-вторых, законы движения, применимые к отдельным частицам, образующим эти тела, как нетрудно видеть, работают и на уровне самих тел, вследствие чего с очень хорошим прибли-

жением Солнце, планеты и луны можно, в свою очередь, рассматривать как частицы и не беспокоиться по поводу малых движений отдельных составляющих небесных тел![10] И снова нам удается свести все к рассмотрению системы из «небольшого» количества тел, где расплывание начальной области в фазовом пространстве становится несущественным.

Помимо небесной механики и поведения запущенных тел (камней, пуль, ядер, и т. д.), что можно рассматривать как ее частный случай, а также изучения простых систем, содержащих небольшое число частиц, — основные методы, использовавшиеся ньютоновской механикой, очевидно, не могут быть вообще отнесены к разряду «детерминистско-предсказуемых» в том смысле, о котором мы говорили выше. Общую ньютоновскую схему используют скорее для построения моделей, изучение которых позволяет делать выводы о поведении системы в целом. Некоторые точные следствия из законов движения, такие, как законы сохранения энергии, импульса и углового момента, действительно выполняются на любых масштабах. Кроме того, существуют статистические свойства, которые можно комбинировать с динамическими законами, управляющими отдельными частицами, и использовать их для общего прогнозирования поведения системы. (См. обсуждение термодинамики в главе 7; эффект расплывания в фазовом пространстве, рассмотрением которого мы занимались выше, находится в достаточно тесной взаимосвязи со вторым началом термодинамики — и при соблюдении надлежащей осторожности эти идеи действительно можно использовать для прогнозирования.) Искусно проделанное самим Ньютоном вычисление скорости звука в воздухе (слегка подправленное столетие спустя Лапласом) — хороший тому пример. Но весьма редко случается, чтобы детерминизм, присущий ньютоновской (или, в более широком смысле, гамильтоновой) динамике, реально использовался на практике.

Эффект расплывания начальной области в фазовом пространстве приводит к еще одному замечательному следствию. Только подумайте: ведь он свидетельствует о том, что *классическая механика, на самом деле, не в состоянии адекватно описать наш с вами мир*! Я несколько преувеличиваю — но не так уж сильно. Классическая механика может достаточно точно описывать поведение жидких тел — главным образом газов, хотя (с приемлемой степенью точности) и собственно жидкостей — в том случае, когда интерес представляют общие «усредненные» свойства систем частиц; но она испытывает затруднения при попытке объяснить структуру твердых тел, которая отличается более высокой организацией. Проблемой здесь становится невозможность описать феномен сохранения твердым телом своей формы несмотря на то, что оно состоит из мириадов точечноподобных частиц, структура относительного расположения которых постоянно нарушается из-за расплывания начальной области в фазовом пространстве. Как мы теперь знаем, для того, чтобы разобраться в строении твердых тел, необходима квантовая теория, поскольку квантовые эффекты могут каким-то образом предотвратить расплывание портрета системы в фазовом пространстве. Это — весьма важный вопрос, к которому мы еще вернемся в дальнейшем (см. главы 8 и 9).

Затронутая нами тема имеет не менее важное значение и для вопроса о построении «вычислительной машины». Эффект расплывания в фазовом пространстве относится к разряду явлений, которые необходимо контролировать. Нельзя позволить слишком сильно расплываться той области фазового пространства, которая соответствует «дискретному» состоянию вычислительного устройства (такой, например, как описанная выше область R_0). Напомним, что даже в «бильярдном компьютере» Фредкина—Тоффоли требовались некоторые специально вводимые извне *твердые стенки*, необходимые для правильной работы компьютера. Объяснить «цельность» объекта, состоящего из множества частиц, можно в действительности только с помощью квантовой механики. Создается впечатление, что даже «классическая» вычислительная машина должна заимствовать некоторые принципы из квантовой физики — иначе она просто не сможет работать эффективно!

Электромагнитная теория Максвелла

В ньютоновской картине мира мы представляем, что крохотные частицы влияют друг на друга с помощью сил, действующих на расстоянии, причем если частицы не совсем точечные, то они способны отскакивать друг от друга в результате прямого физического контакта. Как уже упоминалось раньше (с. 162), электрические и магнитные силы (которые были известны еще с античных времен и впервые подробно изучены Уильямом Гильбертом в 1600 году и Бенджамином Франклином в 1752 году) действуют аналогично гравитационным силам, поскольку также обратно пропорциональны квадрату расстояния — хотя обе представляют собой скорее силы отталкивания, чем притяжения, действуя в соответствии с принципом «подобное отталкивает подобное»; а вместо массы мерой интенсивности их воздействия служит электрический заряд и сила магнитного полюса, соответственно. На этом уровне не существует никаких трудностей, которые препятствовали бы включению электричества и магнетизма в ньютоновскую схему. Поведение света может быть сравнительно легко описано в общем виде с позиций ньютоновской механики (хотя определенные проблемы при этом все же возникают): либо путем рассмотрения света как субстанции, состоящей из отдельных частиц («фотонов», как теперь их принято называть); либо с помощью представления его в виде волнового процесса, распространяющегося в некоторой среде (в последнем случае эту среду — «эфир» — следует считать состоящей из отдельных частиц).

То, что движущиеся электрические заряды могут создавать магнитные силы, вызывает некоторые дополнительные затруднения, но не разрушает целиком всю ньютонианскую схему. Многие математики и физики (в том числе Гаусс) предлагали системы уравнений для описания эффектов, создаваемых движущимися электрическими зарядами. В рамках общей ньютонианской схемы эти уравнения казались вполне удовлетворительными. Первым, кто бросил серьезный вызов «ньютонианской» картине мира, был, по-видимому, великий английский физик-экспериментатор Майкл Фарадей (1791–1867).

Чтобы понять суть этого вызова, необходимо прежде всего разобраться в смысле термина физическое *поле*. Начнем с магнитного поля. Большинству читателей случалось наблюдать за поведением железных опилок, рассыпанных на листке бумаги, который положили поверх магнита. Железные опилки поразительным образом выстраиваются вдоль так называемых «магнитных силовых линий». Представим себе, что силовые линии присутствуют в пространстве, даже если нет железных опилок. Эти силовые линии и образуют то, что мы называем *магнитным полем*. В каждой точке пространства это «поле» ориентировано в определенном направлении, а именно — в направлении силовой линии, проходящей через данную точку. В действительности, мы имеем в каждой точке пространства *вектор*, т. е. магнитное поле является примером векторного поля. (Мы можем сравнить магнитное поле с гамильтоновым векторным полем, которое было рассмотрено нами в предыдущем разделе, но теперь мы имеем векторное поле в обычном, а не фазовом пространстве.) Точно так же и тела, несущие электрический заряд, оказываются окруженными полем, только несколько иного рода, которое известно под названием *электрического поля*; а любое массивное тело создает вокруг себя так называемое *гравитационное поле*. Все это — векторные поля в обычном пространстве.

Подобные идеи были известны задолго до Фарадея, и в ньютоновской механике они составляли весьма заметную часть арсенала теоретиков. Но согласно господствовавшей тогда точке зрения, такие «поля» не рассматривались как реальная физическая субстанция. Их скорее считали своего рода страницами вспомогательной «бухгалтерской книги», в различных точках которых надлежало размещать подходящие частицы. Но фундаментальные явления, наблюдаемые Фарадеем (во время опытов с движущимися витками с током, магнитами и т. п.), привели его к убеждению, что электрическое и магнитное поля совершенно «материальны» с физической точки зрения — и к открытию у переменных полей способности

«проталкивать» друг друга через пустое пространство, порождая своего рода бестелесную волну! Фарадей высказал предположение, что свет может состоять из таких волн. Подобная точка зрения существенно отличалась от господствовавшей в то время «ньютонианской мудрости», которая не считала электромагнитные поля чем-то «реальным», а рассматривала их всего лишь как удобные вспомогательные математические понятия для описания «настоящей» ньютоновской картины «физической реальности» — «действия на расстоянии (дальнодействия) точечных частиц».

Столкнувшись с обнаруженными Фарадеем экспериментальными фактами, а также с более ранними открытиями замечательного французского физика Андре Мари Ампера (1775–1836) и других исследователей, великий шотландский физик и математик Джеймс Клерк Максвелл (1831–1879) задумался над математической формой уравнений, описывающих электрические и магнитные поля с учетом обнаруженных экспериментальных фактов. В результате поразительного интуитивного озарения Максвелл предложил внести в уравнения незначительную на первый взгляд поправку, что привело к поистине фундаментальным последствиям. Эта поправка в принципе не могла быть подсказана ему никакими из известных экспериментальных фактов (хотя и находилась в согласии с ними). Выводы Максвелла были результатом собственных теоретический постулатов Максвелла — отчасти физических, отчасти математических, а где-то — даже эстетических. Одно из следствий уравнений Максвелла говорило о том, что электрическое и магнитное поля действительно «проталкивают» друг друга сквозь пустое пространство. Осциллирующее магнитное поле должно было бы порождать осциллирующее электрическое поле (о чем свидетельствовали экспериментальные факты, полученные Фарадеем); а это осциллирующее электрическое поле, в свою очередь, должно создавать осциллирующее магнитное поле (в согласии с теоретическими выводами Максвелла); последнее снова порождает осциллирующее электрическое поле и т. д. (См. рис. 6.26, 6.27 на с. 240 и 241, где схематически изображен этот волновой процесс.)

Максвеллу удалось вычислить скорость, с которой этот процесс должен был бы распространяться в пространстве, и она в результате оказалась равной скорости света! Кроме того, эти так называемые *электромагнитные* волны *интерферировали* и обладали удивительной способностью поляризоваться, как и свет (последнее свойство на тот момент было уже давно известно, а мы еще вернемся к нему в главе 6, с. 214, 240). Помимо объяснения свойств видимого света, для которого длины электромагнитных волн должны были бы лежать в диапазоне $4-7\times10^{-7}$ м, Максвелл предсказал существование электромагнитных волн других длин, порождаемых электрическими токами в проводниках. Существование таких волн было экспериментально установлено замечательным немецким физиком Генрихом Герцем в 1888 году. Вдохновенная надежда Фарадея воплотилась в чудесные уравнения Максвелла!

Хотя нам совсем не обязательно вдаваться в подробности уравнений Максвелла, давайте все же окинем их быстрым взглядом:

$$\frac{1}{c^2}\times\frac{\partial\mathbf{E}}{\partial t}=\operatorname{rot}\mathbf{B}-4\pi\mathbf{j},\qquad \frac{\partial\mathbf{B}}{\partial t}=-\operatorname{rot}\mathbf{E},$$

$$\operatorname{div}\mathbf{E}=4\pi\rho,\qquad \operatorname{div}\mathbf{B}=0.$$

Здесь $\mathbf{E}$, $\mathbf{B}$ и $\mathbf{j}$ — векторные поля, описывающие, соответственно, электрическое поле, магнитное поле и электрический ток; ρ — плотность электрического заряда, а c — постоянная — скорость света [11]. Не стоит огорчаться, если вам не известен смысл обозначений "rot" и "div". Они просто означают различные пространственные вариации полей $\mathbf{B}$ и $\mathbf{E}$. (Обозначения "rot" и "div" представляют собой определенные комбинации частных производных по пространственным координатам. Напомним, что операции взятия «частной производной», обозначаемой символом ∂, мы коснулись в связи с уравнениями Гамильтона.) Операторы $\partial/\partial t$, стоящие в левых частях двух первых уравнений, по существу означают то же самое, что «точки» в уравнениях Гамильтона (различие в обозначениях вызвано чисто техническими причинами). Таким образом, $\partial\mathbf{E}/\partial t$ означает «скорость изменения во времени электрического поля», а $\partial\mathbf{B}/\partial t$ означает «скорость изменения во времени магнитного поля».

Первое уравнение[10] связывает изменения электрического поля с текущими значениями магнитного поля и электрического тока; тогда как второе, наоборот, описывает изменения магнитного поля в зависимости от величины электрического поля. Третье уравнение, грубо говоря, представляет собой закодированную форму закона обратных квадратов, показывающую, как электрическое поле (в данный момент времени) должно быть связано с распределением зарядов. Что же касается четвертого уравнения, то оно говорит то же самое о магнитном поле (с той лишь разницей, что «магнитные заряды» — отдельные «северные» и «южные» полюсы частиц — не существуют).

Уравнения Максвелла несколько напоминают уравнения Гамильтона тем, что определяют скорость изменения по времени соответствующих величин (электрического и магнитного полей) в зависимости от их текущих значений в любой заданный момент времени. Следовательно, уравнения Максвелла являются по сути *детерминистскими* — точно так же, как и система уравнений в обычной гамильтоновой теории. Единственное (хотя и важное) различие состоит в том, что уравнения Максвелла *полевые*, а не *корпускулярные*. Это означает, что для описания состояния такой системы необходимо *бесконечно много* параметров (векторы поля в каждой точке пространства) вместо всего лишь *конечного* числа параметров (трех координат положения и трех компонент импульса каждой частицы) в корпускулярной теории. Таким образом, фазовое пространство в теории Максвелла *бесконечномерно*! (Как я уже упоминал выше, уравнения Максвелла в действительности могут быть включены в общую гамильтонову схему, но из-за их бесконечномерности гамильтонову схему перед этим необходимо слегка обобщить [12].)

Принципиально *новой* составляющей в той картине нашего физического мира, которая выстраивалась на основе теории Максвелла (помимо и сверх того, что было известно ранее), стала необходимость рассматривать *поля* уже не как математические придатки к «реальным» частицам, или корпускулам, в ньютоновской теории — но как самостоятельно существующие объекты. Действительно, Максвелл показал, что когда поля распространяются в виде электромагнитных волн, они переносят с собой определенное количество *энергии*. Ему удалось получить даже явное выражение для этой энергии. То есть оказалось, что энергию, на самом деле, могли переносить с места на место «нематериальные» электромагнитные волны. Этот факт был экспериментально подтвержден Герцем, сумевшим зарегистрировать электромагнитные волны. То, что радиоволны действительно *могут* переносить энергию, до сих пор представляется удивительным даже тем, кто в той или иной степени знаком с этим феноменом!

Вычислимость и волновое уравнение

Непосредственно из своих уравнений Максвелл сумел вывести, что в областях пространства, где нет ни зарядов, ни токов (т. е. там, где в приведенных выше уравнениях $\mathbf{j} = 0$, $\rho = 0$) все компоненты электрического и магнитного полей должны удовлетворять так называемому *волновому* уравнению[11]. Волновое уравнение можно рассматривать как «упрощенный вариант» уравнений Максвелла, так как оно записано для *одной-единственной* величины, а не для всех шести компонент электрического и магнитного полей. Решения уравнения Даламбера дают пример волнообразного движения без дополнительных усложняющих свойств наподобие «поляризации» в теории Максвелла (направления вектора электрического поля, см. с. 240).

Волновое уравнение представляет сейчас для нас тем больший интерес, что оно

---

[10] Именно введение $\partial\mathbf{B}/\partial t$ в это уравнение было мастерским штрихом в теоретических рассуждениях Максвелла. Все остальные члены во всех уравнениях, по существу, были известны из опытных данных. Что же касается коэффициента $1/c^2$, то он очень мал и поэтому член с $\partial\mathbf{B}/\partial t$ не мог быть обнаружен экспериментально.

[11] Волновое уравнение (уравнение Даламбера) представимо в виде

$$\left\{\left(\frac{1}{c^2}\right)\times\left(\frac{\partial}{\partial t}\right)^2-\left(\frac{\partial}{\partial x}\right)^2-\left(\frac{\partial}{\partial y}\right)^2-\left(\frac{\partial}{\partial z}\right)^2\right\}\varphi=0.$$

было предметом целенаправленного изучения именно в связи с его свойствами *вычислимости*. Действительно, Мариану Б. Пур-Элю и Яну Ричардсу (Пур-Эль, Ричардс [1979, 1981, 1982], см. также [1989]) удалось показать, что, даже несмотря на *детерминистское* (в обычном смысле) поведение решения волнового уравнения — при котором данные в начальный момент времени однозначно определяют решение во все остальные моменты времени — существуют *вычислимые* начальные данные некоего «особого» рода, обладающие тем свойством, что для них однозначно рассчитать значения поля в более поздний (вычислимый) момент времени — *невозможно*. Таким образом, уравнения вполне допустимой физической теории поля (хотя и отличающейся от теории Максвелла, которая действительно «работает» в нашем мире) могут, согласно Пур-Элю и Ричардсу, породить невычислимую эволюцию!

На первый взгляд это кажется весьма удивительным результатом, который вроде бы противоречит тому, о чем я говорил в предыдущем разделе относительно возможной вычислимости «разумных» гамильтоновых систем. Однако, несмотря на то, что поразительный результат Пур-Эля—Ричардса исполнен, несомненно, глубокого *математического* смысла, он все же не противоречит высказанной выше гипотезе — причем по причине, имеющей глубокий *физический* смысл. Причина же эта состоит в том, что начальные условия «особого» рода не относятся к «плавно изменяющимся» [13], а именно это свойство обычно требуется от каждого поля, имеющего физический смысл. Пур-Эль и Ричардс в действительности доказали, что невычислимость *не может возникнуть* в случае волнового уравнения, если мы не будем рассматривать поля «особого» рода. С другой стороны, даже если бы такие поля считались допустимыми, было бы трудно понять, как может использовать подобную «невычислимость» любое физическое «устройство» (например, головной мозг человека)? Она могла бы иметь существенное значение только при наличии возможности производить измерения со сколь угодно высокой степенью точности (которые, как я объяснял выше, нереальны с физической точки зрения). Тем не менее, результаты Пур-Эля—Ричардса открывают интригующую область знания, которая до сих пор остается практически нетронутой.

Уравнение движения Лоренца; убегающие частицы

Система уравнений Максвелла в том виде, как мы ее выписали, не является, на деле, полной. Эти уравнения великолепным образом описывают распространение электрических и магнитных полей при наличии *заданного* распределения электрических зарядов и токов. Эти заряды физически нам даны в виде *заряженных частиц* — в основном, электронов и протонов, как нам сейчас известно — а токи порождаются движением этих частиц. Если мы знаем, где находятся заряженные частицы и как они движутся, то уравнения Максвелла позволяют определить поведение электромагнитного поля. Но вот что уравнения Максвелла нам *не говорят* — это как должны себя вести сами частицы. Частичный ответ на этот вопрос был известен еще во времена Максвелла, но удовлетворительной системы уравнений не было до тех пор, пока в 1895 году замечательный голландский физик Хендрик Антон Лоренц, воспользовавшись идеями, близкими к идеям специальной теории относительности, не вывел уравнения движения заряженной частицы, известные ныне как *уравнения Лоренца* (см. Уиттекер [1910]). Эти уравнения позволяют описывать непрерывные изменения скорости заряженной частицы под действием электрического и магнитного полей в той точке, где она в данный момент находится [14]. Присоединив уравнения Лоренца к уравнениям Максвелла, мы получаем систему уравнений, описывающих эволюцию во времени *и* заряженных частиц, *и* электромагнитного поля.

Но эта система уравнений, в свою очередь, тоже не безукоризненна. Она дает превосходные результаты, если поля однородны вплоть до масштабов порядка диаметра самих частиц (за единицу измерения диаметра принимается «классический радиус»

электрона — около 10^{-15} м), а движения частиц не слишком интенсивны. Однако здесь имеется принципиальная трудность, обойти которую при других обстоятельствах становится невозможно. Дело в том, что уравнения Лоренца подразумевают измерения электромагнитного поля в той самой точке, где находится заряженная частица (по существу, такое измерение должно дать нам значение «силы», действующей в этой точке со стороны электромагнитного поля на нашу частицу). Но где следует выбирать эту точку, если частица имеет конечные размеры? Следует ли принять за нужную точку «центр» частицы, или поле («силу») необходимо усреднить по всем точкам поверхности частицы? Если поле неоднородно в масштабе порядка размера частицы, то разный выбор точки может привести к отличающимся результатам. Есть и другая, более серьезная проблема: каково *на самом деле* электромагнитное поле на поверхности частицы (или в ее центре)? Напомним, что мы рассматриваем *заряженную* частицу. Следовательно, электромагнитное поле, *обусловленное самой* частицей, необходимо добавить к «фоновому полю», в котором находится частица. Вблизи самой «поверхности» частицы ее собственное поле становится чрезвычайно интенсивным и легко поглощает все остальные поля в окрестности частицы. Кроме того, собственное поле частицы всюду вокруг нее направлено преимущественно наружу (или вовнутрь), вследствие чего результирующее *истинное* поле, на которое по предположению реагирует частица, вовсе не однородно, а в каждой точке на «поверхности» частицы направлено в свою сторону, не говоря уже о «внутренности» частицы (рис. 5.15). Дополнительно к этому нам следует выяснить, будут ли отличающиеся по величине силы, которые действуют на частицу, стремиться повернуть или деформировать ее; а также понять, какими упругими свойствами обладает частица: и т. д. (особенно трудны вопросы возникающие в связи с *теорией относительности*, но я не собираюсь сейчас отвлекать на них внимание читателя). Ясно, что теперь проблема становится намного сложнее по сравнению с тем, какой она казалась нам прежде.

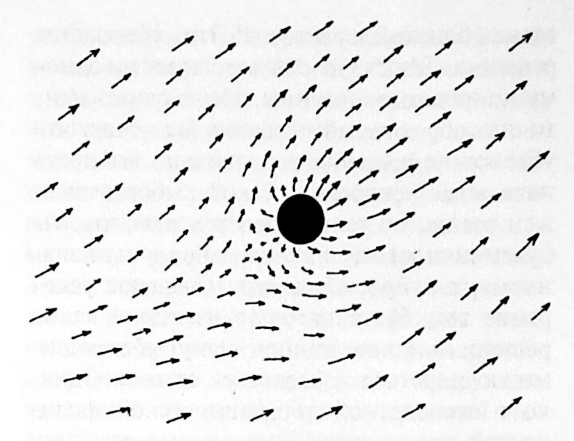


Рис. 5.15. Как можно строго применить уравнения движения Лоренца? Сила, действующая на заряженную частицу, не может быть получена измерением поля в точке нахождения частицы, так как здесь доминирует собственное поле частицы

Возможно, нам стоило бы рассматривать частицу как *материальную точку*. Но такой подход приводит к проблемам другого рода, ибо в непосредственной окрестности точечной частицы ее собственное электрическое поле становится *бесконечным*. Если, как это следует из уравнений Лоренца, частица должна реагировать на электромагнитное поле в той точке, где она находится, то точечная частица должна испытывать действие со стороны бесконечно большого поля! Чтобы формула Лоренца для величины силы имела смысл, необходимо найти способ, который позволил бы *вычитать собственное* поле частицы и оставлять конечное фоновое поле, которое бы однозначно определяло поведение частицы. Такой метод был предложен в 1938 году Дираком (о котором мы еще услышим в дальнейшем). Однако решение Дирака приводило к определенным следствиям, которые не могли не вызывать тревогу. Дирак обнаружил, что для однозначного определения поведения частиц и полей исходя из соответствующих начальных данных, необходимо знать не только начальное положение и скорость каждой частицы, но и ее начальное *ускорение* (в контексте стандартных динамических теорий такую ситуацию нельзя не признать несколько аномальной). Для большинства значений начального ускорения частица ведет себя самым «сумасшедшим» образом, спонтанно ускоряясь в пространстве до скорости,

весьма близкой к световой! Эти «убегающие решения» Дирака не соответствуют ни одному природному явлению. Необходимо найти способ, который позволил бы исключать убегающие решения и правильно выбирать начальные ускорения. Такой выбор возможен всегда, но только при условии, что мы будем пользоваться неким «априорным знанием», т. е. будем задавать начальное ускорение так, будто нам уже известно, какие решения в конце концов станут убегающими, и стараться избавляться от них. Однако в стандартной детерминистской физической задаче начальные данные задаются по-другому — произвольно и без каких-либо ограничений и требований относительно будущего поведения решений. В нашем же случае не только будущее полностью определено данными, заданными в некоторый момент времени в прошлом, но и сам способ задания этих данных весьма жестко ограничен требованием, накладываемым на будущее («допустимое») поведение частиц и полей!

Так обстоит дело, пока мы рассматриваем фундаментальные классические уравнения. Читатель легко поймет, что вопрос о детерминизме и вычислимости в законах классической физики носит раздражающе неясный характер. Действительно ли в физических законах есть *телеологическая* составляющая, которая заставляет будущее каким-то образом оказывать влияние на происходящее в прошлом? На самом деле, физики обычно не рассматривают подобные следствия из *классической электродинамики* (теории классических заряженных частиц, а также электрического и магнитного полей) как соответствующие реальности. Стандартный ответ физиков на упомянутые выше трудности сводится к утверждению, что «отдельные заряженные частицы» относятся к области *квантовой электродинамике*, и что нельзя ожидать получить разумные ответы на подобные вопросы, если использовать строго классическую теорию. Такое утверждение, безусловно, верно — но, как мы увидим в дальнейшем, в *самой* квантовой теории здесь также возникают проблемы. На самом деле, Дирак исследовал классическую задачу движения заряженной частицы именно *потому*, что надеялся обнаружить там какие-нибудь новые идеи, способные помочь в разрешении еще более фундаментальных трудностей, возникающих при рассмотрении (физически более адекватной) квантовой задачи. С проблемами квантовой теории нам еще придется столкнуться позднее!

Специальная теория относительности Эйнштейна и Пуанкаре

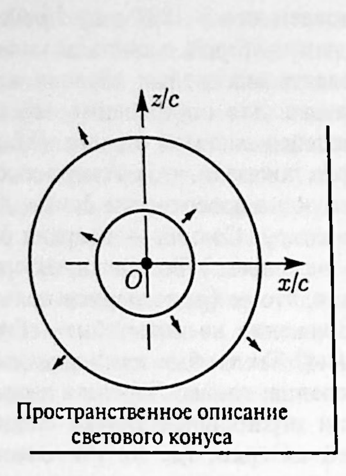
Напомним принцип относительности Галилея, который гласит, что физические законы Ньютона и Галилея останутся совершенно неизменными, если от покоящейся системы отсчета мы перейдем в другую, движущуюся равномерно и прямолинейно. Из этого принципа следует, что, просто наблюдая динамическое поведение объектов вокруг нас, мы не можем установить, находимся ли мы в состоянии покоя или движемся равномерно и прямолинейно в каком-то направлении. (Вспомним пример Галилея с кораблем в море, см. с. 158.) Предположим теперь, что к законам Ньютона и Галилея мы присоединили уравнения Максвелла. Останется ли при этом в силе принцип относительности Галилея? Напомним, что электромагнитные волны Максвелла распространяются с фиксированной скоростью c — скоростью света. Казалось бы, здравый смысл подсказывает, что если мы будем двигаться очень быстро в каком-нибудь направлении, то должно создаться впечатление, что скорость света в этом направлении *стала меньше c* (поскольку мы «догоняем свет»); а скорость света в противоположном направлении — больше c (так как при этом мы движемся «от света»). Видно, что и тот, и другой результат отличны от *фиксированного* значения c скорости света в теории Максвелла. И здравый смысл не подвел бы нас: комбинация уравнений Ньютона и Максвелла *не удовлетворяет* принципу относительности Галилея.

Обеспокоенность этими проблемами привела Эйнштейна в 1905 году — а Пуанкаре даже несколько раньше (в 1898–1905 годах), — к созданию специальной теории относительности. Пуанкаре и Эйн-

штейн независимо обнаружили, что уравнения Максвелла *тоже* удовлетворяют некоторому принципу относительности (см. Пайс [1982]), т. е. остаются неизменными при переходе от неподвижной системы отсчета к движущейся, хотя правила такого перехода несовместимы с физикой Галилея—Ньютона! Чтобы сделать их совместимыми, необходимо видоизменить либо одну, либо другую систему уравнений — или же отказаться от принципа относительности. Эйнштейн не собирался отказываться от принципа относительности. Его великолепная физическая интуиция настойчиво подсказывала ему, что принцип относительности должен выполняться для физических законов нашего мира. Кроме того, Эйнштейну было хорошо известно, что практически для всех известных явлений физика Галилея—Ньютона была экспериментально проверена только при скоростях ничтожно малых по сравнению со скоростью света, при которых отмеченная выше несовместимость была несущественной. Только *сам свет*, как было известно, способен развивать скорости достаточно большие для того, чтобы упомянутые выше «несоответствия» начинали заметно сказываться. Следовательно, именно поведение света могло бы подсказать, какой принцип относительности следует избрать; при этом уравнения, которыми описывается свет — это уравнения Максвелла. Таким образом, выбор следовало бы остановить на принципе относительности для теории Максвелла, а законы Галилея—Ньютона — соответственно, модифицировать!

Лоренц еще до Пуанкаре и Эйнштейна тоже заинтересовался этими вопросами и даже нашел на них частичные ответы. К 1895 году Лоренц пришел к заключению, что силы, связывающие частицы материи, имеют электромагнитную природу (как в действительности и оказалось), поэтому поведение реальные материальных тел должно удовлетворять законам, вытекающим из уравнений Максвелла. Отсюда, в частности, следовало, что тело, движущееся со скоростью, сравнимой со скоростью света, должно претерпевать небольшое сокращение в направлении движения («сокращение Фитцджеральда—Лоренца»). Это вывод Лоренц использовал для объяснения удивительного экспериментального факта, установленного в 1897 году Майкельсоном и Морли, который свидетельствовал о том, что электромагнитные явления нельзя использовать для определения «абсолютной» покоящейся системы отсчета. (Майкельсон и Морли показали, что на скорость света, измеряемую на поверхности Земли, движение Земли вокруг Солнца — вопреки ожиданиям — не влияет.) Всегда ли материя ведет себя так, что ее (равномерное прямолинейное) движение не может быть обнаружено локально? Таким был *предварительный* вывод Лоренца; однако Лоренц в своем исследовании ограничился только специальной теорией материи, где не учитывались никакие силы, кроме электромагнитных. Пуанкаре, будучи выдающимся математиком, сумел в 1905 году *строго* показать, что материя должна вести себя именно так, как предполагал в своих теоретических построениях Лоренц — т. е. в соответствии с принципом относительности, лежащим в основе уравнений Максвелла — поэтому равномерное прямолинейное движение вообще не может быть обнаружено локально. Ему также удалось глубоко понять физические следствия из этого принципа (в том числе — явление «относительности одновременности», о котором мы еще поговорим далее). По-видимому, Пуанкаре рассматривал этот принцип лишь как *одну* из возможностей, и не разделял убеждения Эйнштейна, что определенный принцип относительности *должен* выполняться.

Принцип относительности, которому удовлетворяют уравнения Максвелла, ставший известным под названием *специальной* (*или частной*) *относительности* (*СТО*), довольно труден для понимания и полон противоречащих нашей интуиции моментов, которые на первый взгляд невозможно связать с реальными свойствами окружающего нас мира. Действительно, принципу специальной относительности вряд ли удастся придать смысл, если не воспользоваться *еще одной* идеей, введенной в 1908 году немецким геометром русского происхождения Германом Минковским (1864–1909), обладавшим в высшей степени незаурядным мышлением и тонкой интуицией. Минковский был одним из преподавателей Эйн-



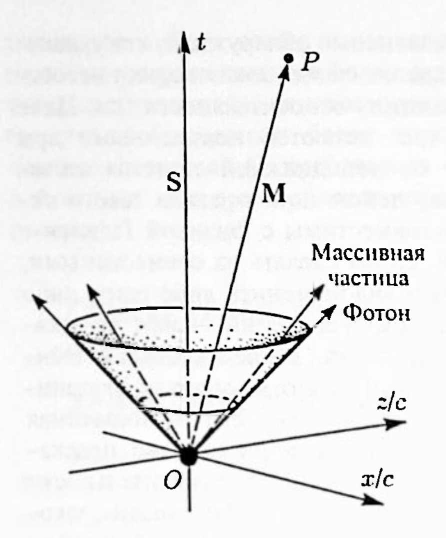


Рис. 5.16. Световой конус в пространстве-времени Минковского (с двумя пространственными измерениями), описывающий историю световой вспышки при взрыве, произошедшем в точке O пространства-времени

штейна в Цюрихском Высшем Политехническом Училище. Его принципиально новая идея состояла в том, что пространство и время следует рассматривать совместно как единую сущность — *четырехмерное пространство-время*. В своей знаменитой лекции, прочитанной в 1908 году в Геттингенском университете, Минковский провозгласил:

> «Таким образом, пространство само по себе и время само по себе обречены исчезнуть, превратившись в бесплотные тени, и только объединение пространства и времени сохранится как независимая реальность».

Попытаемся понять основные положения специальной теории относительности в терминах величественного пространства-времени Минковского.

Одна из трудностей на пути к освоению понятия пространства-времени связана с его *четырехмерностью*, мешающей нам представить себе пространство-время наглядно. Но после того, как мы пережили нашу встречу с фазовым пространством, представить себе всего лишь четыре измерения не составит для нас особых трудностей! Как и в случае с фазовым пространством, мы пойдем на «обман» и нарисуем картину пространства меньшего числа измерений, но теперь степень обмана будет несравненно меньше, а картина, соответственно — гораздо точнее и ближе к истинной. Для многих целей достаточно рассмотреть двумерное пространство-время (одно измерение — для пространства и одно измерение — для времени). Я надеюсь, что читатель простит мне некоторую напористость и разрешит подняться до трехмерного пространства-времени (два измерения — для пространства, и одно измерение — для времени). Это позволит нарисовать вполне убедительную картину, посмотрев на которую нетрудно будет понять, что, в принципе, аналогичные идеи могут быть легко распространены без особых изменений и на четырехмерный случай. Рассматривая графическое изображение пространства-времени, необходимо иметь в виду, что каждая точка на картинке представляет некоторое *событие*, т. е. определенную точку в пространстве в какой-то конкретный момент времени. Иначе говоря, точка пространства-времени обладает только *мгновенным* существованием. Полная картина пространства-времени изображает всю историю: прошлое, настоящее и будущее. Любая частица, коль скоро она существует на протяжении некоторого времени, представляется в пространстве-времени не точкой, а линией, которая называется *мировой линией* данной частицы. Мировая линия — прямая в случае равномерного движения частицы, и искривленная,

если частица движется с ускорением (т. е. *неравномерно*) — описывает всю историю существования частицы.

На рис. 5.16 я изобразил пространство-время с двумя пространственными измерениями и одним временным. Можно считать, что существует обычная временная координата t, измеряемая по вертикали, и две пространственные координаты x/c и z/c, измеряемые по горизонтали [12]. Конус с вершиной в центре — это *световой* конус (будущего), с центром в начале координат O пространства-времени. Чтобы по достоинству оценить его значение, представьте себе, что в точке O происходит взрыв. (Иначе говоря, взрыв происходит в начале пространства в момент времени $t = 0$.) Этот световой конус описывает историю света, испущенного при взрыве. На языке двумерного пространства история вспышки света была бы окружностью, расширяющейся со скоростью света c. В полном трехмерном пространстве вместо окружности мы имели бы *сферу*, расширяющуюся со скоростью света c, — сферический волновой фронт света. Но в рассматриваемом примере мы «подавляем» пространственное направление y и поэтому получаем всего лишь окружность, подобную круговым волнам, расходящимся от точки падения камня на поверхность пруда. Нетрудно понять, что на объемной картине пространства-времени мы получим расширяющиеся окружности, если рассмотрим серию горизонтальных сечений светового конуса, каждое последующее из которых расположено выше предыдущего. Эти горизонтальные сечения представляют собой различные пространственные описания волнового фронта света по мере возрастания временной координаты t. Одна из отличительных особенностей специальной теории относительности состоит в том, что никакая материальная частица не может двигаться быстрее света (подробнее об этом — чуть позднее). Все материальные частицы, возникшие при взрыве, должны отставать от света. На языке пространства-времени это означает, что мировые линии всех частиц, испущенных при взрыве, должны лежать *внутри* светового конуса.

Часто свет бывает удобно описывать не электромагнитными волнами, а как поток частиц, называемых *фотонами*. Мы можем мысленно представлять себе «фотон» как крохотный «пакет» электромагнитного поля, осциллирующего с высокой частотой. Термин «волновой пакет» физически более приемлем в контексте квантовых описаний, к которым мы перейдем в следующей главе, но пока для нас будут полезны и «классические» фотоны. В свободном пространстве фотоны всегда движутся по прямолинейным траекториям с постоянной скоростью c. Это означает, что, изображенная на картине пространства-времени Минковского мировая линия фотона всегда имеет вид прямой, образующей с вертикалью угол 45°. Фотоны, образовавшиеся при взрыве в точке O пространства-времени, описывают световой конус с вершиной в O.

Описанными выше свойствами должны обладать все точки пространства-времени. В начале пространства-времени нет ничего особенного: точка O ничем не отличается от любой другой точки. Следовательно, в любой точке пространства-времени должен быть свой световой конус, имеющий такой же смысл, как и световой конус, исходящий из начала пространства-времени. История любой вспышки света, или мировые линии фотонов, если угодно воспользоваться корпускулярным описанием света, всегда располагаются на поверхности светового конуса с вершиной в каждой точке пространства-времени — тогда как история любой материальной частицы всегда должна располагаться внутри соответствующего светового конуса. Это показано на рис. 5.17. Семейство световых конусов во всех точках пространства-времени можно рассматривать как часть *геометрии Минковского* пространства-времени.

Что такое геометрия Минковского? Самая важная ее часть — структура светового конуса, хотя геометрия Минковского ею не исчерпывается. В этой геометрии существует понятие «расстояния», во многом аналогичное определению расстояния в евклидовой геометрии. В трехмерной евклидо-

---

[12] Причина, по которой пространственные координаты мы делим на c (скорость света), проста: это делается для того, чтобы мировые линии фотонов были наклонены под удобным углом 45° к вертикали (см. текст далее).

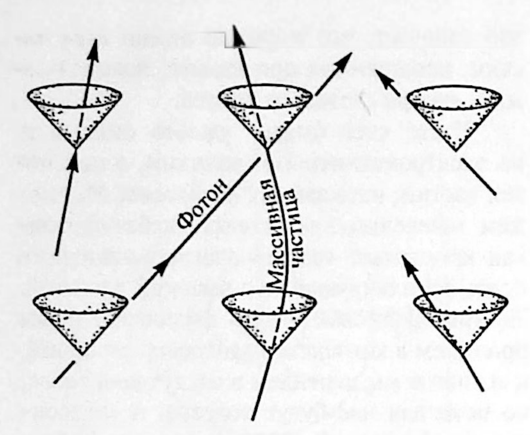


Рис. 5.17. Картина геометрии Минковского

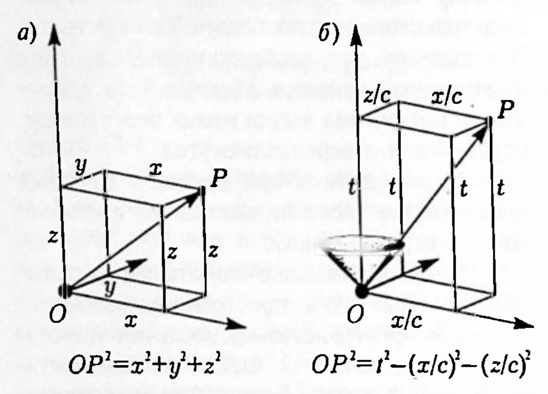


Рис. 5.18. Сравнение «расстояний», измеренных в (*а*) евклидовой геометрии и (*б*) геометрии Минковского (здесь «расстояние» означает «прожитое время»)

вой геометрии расстояние r от произвольной точки до начала координат, выраженное через обычные декартовы координаты, определяется соотношением

$$r^2 = x^2 + y^2 + z^2.$$

(См. рис. 5.18 *а*. Это — всего лишь теорема Пифагора; возможно, двумерный вариант этого соотношения более привычен читателю.) В нашей трехмерной геометрии Минковского выражение для расстояния очень похоже на евклидово (рис. 5.18 *б*); существенное отличие состоит в том, что в геометрии Минковского это выражение содержит два *знака минус*:

$$s^2 = t^2 - \left(\frac{x}{c}\right)^2 - \left(\frac{z}{c}\right)^2.$$

Каков физический смысл величины «расстояния» s в этом выражении? Предположим, что мы рассматриваем точку P с координатами $(t, x/c, y/c, z/c)$, или $(t, x/c, z/c)$ в трехмерном случае; см. рис. 5.16 — она лежит в световом конусе (будущего) точки O. Тогда прямолинейный отрезок OP может представлять часть истории какой-то материальной частицы, например, испущенной при взрыве. «Длина» Минковского s отрезка OP допускает прямую физическую интерпретацию. Это — продолжительность (длина) *интервала времени*, реально прожитого частицей между событиями O и P! Иначе говоря, если бы существовали очень прочные и точные часы, намертво прикрепленные к частице [15], то разность между их показаниями в точках O и P составила бы ровно s единиц времени. Вопреки ожиданиям, величина t сама по себе *не* описывает время, измеряемое этими гипотетическими часами — за исключением того случая, когда часы «покоятся» в нашей системе координат (т. е. имеют фиксированные значения координат $x/c, y/c, z/c$), а это означает, что мировая линия часов имеет на «картине» вид вертикальной прямой. Таким образом, t будет задавать «время» только для тех наблюдателей, которые «стационарны» (т. е. чьи мировые линии — «вертикальные» прямые). *Правильной* мерой времени для движущегося (равномерно и прямолинейно из начала координат O) наблюдателя, согласно специальной теории относительности, служит величина s. Заключение, к которому мы пришли, весьма удивительно и полностью расходится с находящейся в согласии со «здравым смыслом» галилеево-ньютонианской мерой времени, которая просто совпадает с координатным значением t. Обратите внимание на то, что релятивистская (в смысле Минковского) мера времени s всегда несколько *меньше*, чем t, если вообще существует какое-то движение (так как s^2 меньше, чем t^2, коль скоро не все координаты $x/c, y/c, z/c$ равны нулю), как это следует из приведенной выше формулы. Наличие движения (т. е. случай, когда отрезок OP расположен не вдоль оси t) приводит к «замедлению» хода часов по сравнению с t, иными словами, по отношению к показаниям часов в нашей системе отсчета. Если скорость движения мала по сравнению с c, то величины s и t почти совпадают,

чем объясняется то, что мы непосредственно не ощущаем «замедление хода движущихся часов». В другом предельном случае, когда скорость движения совпадает со скоростью света, точка P лежит *на* световом конусе, и мы получаем $s = 0$. Световой конус есть не что иное, как геометрическое место точек, для которых «расстояние» в смысле Минковского (т. е. «время») от начала координат O действительно равно нулю. Таким образом, фотон вообще «не ощущает», как течет время! (Мы не можем позволить себе рассматривать *еще более* экстремальный случай, когда точка P движется у самой поверхности снаружи светового конуса, так как это привело бы к мнимому значению s — квадратному корню из отрицательного числа, и нарушило бы правило, согласно которому материальные частицы, или фотоны, не могут двигаться быстрее света.) [13)]

Понятие «расстояния» в смысле Минковского одинаково хорошо применимо к *любой* паре точек в пространстве-времени, одна из которых лежит внутри светового конуса другой, так что частица может двигаться из одной точки в другую. Мы просто будем считать, что начало координат O перенесено в какую-то иную точку пространства-времени. Кроме того, расстояние по Минковскому между точками соответствует интервалу времени, отсчитываемого часами, которые равномерно и прямолинейно движутся из одной точки в другую. Когда в качестве частицы выступает фотон, и расстояние в смысле Минковского обращается в нуль, мы получаем две точки, одна из которых лежит *на* световом конусе другой — что позволяет *строить* световой конус для последней.

Основная структура геометрии Минковского со столь причудливой мерой «длины» мировых линий, интерпретируемой как *время*, измеряемое (или «прожитое») физическими часами, несет в себе самую суть специальной теории относительности. В частности, читателю, возможно, известен так называемый «парадокс близнецов» в СТО:

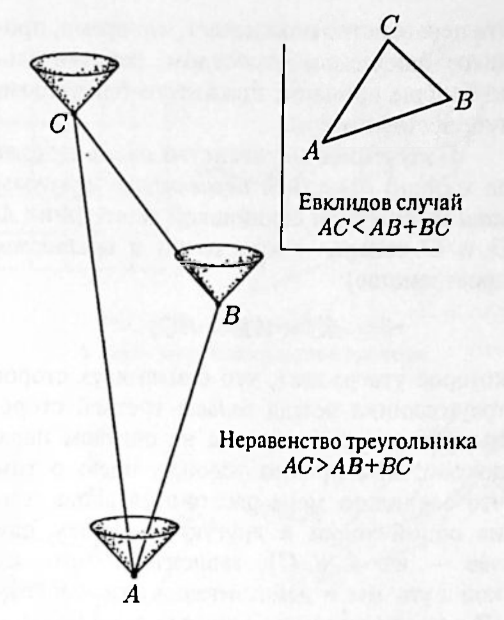


Рис. 5.19. Так называемый «парадокс близнецов» специальной теории относительности, трактуемый с помощью неравенства треугольника в геометрии Минковского. (Для сравнения приведен и евклидов случай.)

один из братьев-близнецов остается на Земле, другой совершает путешествие на соседнюю звезду, двигаясь туда и обратно с огромной скоростью, приближающейся к скорости света. По возвращении выясняется, что близнецы состарились неодинаково: путешественник все еще молод, а его брат, остававшийся на Земле, стал дряхлым стариком. «Парадокс близнецов» легко описывается в терминах геометрии Минковского, и всякий может без труда понять, почему это явление — хотя и способное озадачить — парадоксальным все же не является. Мировая линия AC принадлежит тому из близнецов, который остается дома, тогда как мировая линия близнеца-путешественника состоит из двух отрезков AB и BC, соответствующих полету на звезду и возвращению на Землю (рис. 5.19). Близнец-домосед проживает время, измеряемое расстоянием в смысле Минковского AC, тогда как близнец-путешественник проживает время, измеряемое суммой [16] двух расстояний AB и BC. Эти времена не равны, и мы обнаруживаем, что

$$AC > AB + BC.$$

13) Тем не менее для событий, разделенных отрицательными значениями s^2, величина $c^2\sqrt{-s^2}$ имеет смысл, равняясь обычному расстоянию до того наблюдателя, которому события кажутся одновременными (см. далее).

Это неравенство показывает, что время, прожитое близнецом-домоседом, действительно больше времени, прожитого близнецом-путешественником.

Полученное неравенство очень похоже на хорошо известное *неравенство треугольника* из обычной евклидовой геометрии (A, B и C теперь — три точки в *евклидовом* пространстве):

$$AC < AB + BC,$$

которое утверждает, что сумма двух сторон треугольника всегда *больше* третьей стороны. Это неравенство мы не считаем парадоксом! Мы прочно усвоили идею о том, что евклидова мера расстояния вдоль пути из одной точки в другую (в нашем случае — из A в C), зависит от того, какой путь мы в действительности выберем. (В рассматриваемом примере двумя путями служат AC и более длинный изломанный маршрут ABC.) Неравенство треугольника — частный случай общего утверждения, которое гласит, что кратчайшее расстояние между двумя точками (в данном случае A и C) измеряется по прямой, их соединяющей (отрезок AC). Изменение знака неравенства на обратный при измерении расстояний в смысле Минковского происходит вследствие изменения знаков в определении «расстояния», в результате чего отрезок AC, измеряемый по Минковскому, оказывается «длиннее», чем ломаный маршрут ABC. Таким образом, «неравенство треугольника» в геометрии Минковского в более обобщенной формулировке говорит о том, что *самой длинной* (в смысле наибольшего прожитого времени) среди мировых линий, соединяющих два события, является прямая (т. е. траектория, соответствующая равномерному движению). Если оба близнеца стартуют из точки A и завершают свой путь в точке C, и при этом первый близнец движется прямо из A в C без ускорения, а второй — с ускорением, то первый близнец к моменту встречи со вторым всегда успевает прожить более длинный интервал времени.

Может показаться возмутительным вводить столь странную и сильно расходящуюся с нашими интуитивными представлениями концепцию меры времени. Однако ныне имеется огромное количество экспериментальных данных, свидетельствующих о правомерности такого положения. Например, существует много субатомных частиц, которые распадаются (т. е. превращаются в другие частицы) в определенной шкале времени. Иногда такие частицы движутся со скоростями, очень близкими к скорости света (например, так происходит с космическими лучами, попадающими на Землю из космического пространства, или в созданных человеком ускорителях элементарных частиц), и их времена распада оказываются при этом «растянуты» в полном согласии с вышеизложенными рассуждениями. Еще удивительнее другое: теперь, когда стало возможным изготовить особо точные («ядерные») часы, мы можем *непосредственно* обнаружить эффекты замедления хода часов, перевозимых на высокоскоростных самолетах, летающих на небольшой высоте — причем результаты измерений согласуются с мерой «расстояния» s в смысле Минковского, а не с t! (Строго говоря, с учетом *высоты* приходится принимать во внимание небольшие дополнительные гравитационные эффекты, предсказываемые *общей* теорией относительности, но они также согласуются с наблюдениями — см. следующий раздел.) Кроме того, существует много других явлений, тесно связанных со всей теоретической основой СТО, постоянно подтверждающейся вплоть до мельчайших деталей. Одно из них — знаменитое соотношение Эйнштейна

$$E = mc^2,$$

которое по существу устанавливает равноправие энергии и массы. (В конце этой главы мы познакомимся с некоторыми необычайно заманчивыми следствиями из этого соотношения.)

Я еще не объяснил, каким образом принцип относительности оказывается реально включенным в намеченную выше схему. Каким образом происходит, что наблюдатели, движущиеся прямолинейно и равномерно с различными скоростями, могут оказаться *эквивалентными* с точки зрения геометрии Минковского? Каким образом ось времени на рис. 5.16 («стационарный наблюдатель») может быть полностью эквивалентной некоторой другой прямолинейной мировой линии, например, отрезку OP

(«движущийся наблюдатель»)? Задумаемся сначала над особенностями *евклидовой* геометрии. Ясно, что в ней две произвольные несовпадающие прямые совершенно эквивалентны по отношению к геометрии в целом. Можно мысленно представить себе, что все евклидово пространство «скользит» по самому себе как «жестко скрепленное целое» до тех пор, пока одна прямая не совпадет с другой. Представьте себе *двумерный* случай — евклидову *плоскость*. Можно представить себе листок бумаги, жестко скользящий по плоской поверхности, до тех пор, пока некоторая прямая, проведенная на листке бумаги, не совпадет с прямой, проведенной на поверхности. Это «жесткое» движение сохраняет структуру геометрии. Аналогичное утверждение справедливо и относительно геометрии Минковского, хотя это и менее очевидно, так что следует проявлять особую осмотрительность, договариваясь о том, какой смысл надлежит вкладывать в термин «жесткое движение». Вместо листка бумаги следует рассматривать особый материал (возьмем сначала для простоты двумерный случай), на котором прямые с углом наклона 45° сохраняют этот угол, тогда как сам материал может растянуться в одном направлении под углом 45° и, соответственно, сжаться в другом направлении под углом 45°. Такая ситуация изображена на рис. 5.20. На рис. 5.21 я попытался показать, что происходит в трехмерном случае. Эта разновидность «жесткого движения» пространства Минковского, называемая *движением Пуанкаре* (или неоднородным движением Лоренца), может выглядеть не очень «жесткой», но она сохраняет все расстояния в смысле Минковского, а «сохранение всех расстояний» — это ни что иное, как смысл понятия «жесткий» в евклидовом случае. Принцип специальной относительности утверждает, что законы физики при таких движениях Пуанкаре пространства-времени остаются неизменными. В частности, «стационарный» наблюдатель S, мировая линия которого совпадает с осью времени на нашем исходном изображении пространства-времени Минковского (рис. 5.16), имеет дело с физикой, совершенно эквивалентной физике

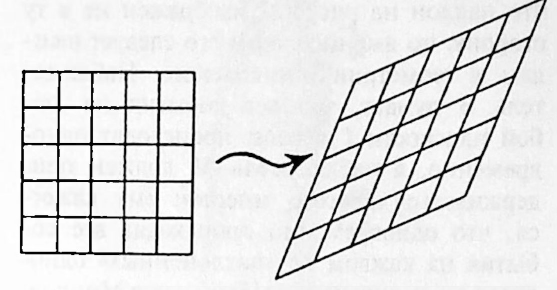


Рис. 5.20. Движение Пуанкаре в двумерном пространстве-времени

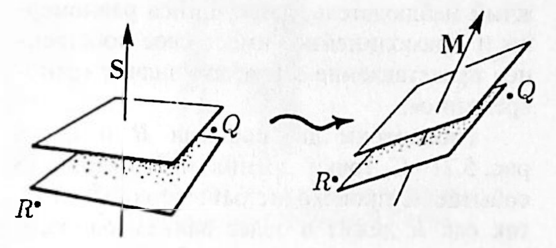


Рис. 5.21. Движение Пуанкаре в трехмерном пространстве-времени. На рисунке справа изображены пространства, одновременные для наблюдателя S, на рисунке слева — одновременные для наблюдателя M. Обратите внимание, что, по мнению наблюдателя S, событие R предшествует событию Q, тогда как, с точки зрения наблюдателя M, событие Q предшествует событию R. (Движение в данном случае считается *пассивным*, т. е. приводит лишь к различным *описаниям* двумя наблюдателями S и M одного и того же пространства-времени.)

«движущегося» наблюдателя M с мировой линией вдоль прямой OP.

Каждая координатная плоскость $t = \text{const}$ представляет для наблюдателя S «пространство» в какой-то один момент «времени», т. е. семейство событий, которые он считает *одновременными* (происходящими в «одно и то же время»). Назовем эти плоскости *одновременными пространствами* наблюдателя S. Когда же мы переходим к другому наблюдателю M, то с необходимостью переводим наше исходное семейство одновременных пространств в некоторое новое семейство с помощью движения Пуанкаре, что позволяет нам получить одновременные пространства для наблюдателя M [17]. Обратите внимание на то, что одновременные пространства наблюдателя M выглядят «наклоненными вверх» (рис. 5.21). Если мыслить в терминах жестких движений в евклидовой геометрии, то может показаться,

что наклон на рис. 5.21 изображен не в ту сторону, но именно таким его следует ожидать в геометрии Минковского. Наблюдатель **S** думает, что все события на любой плоскости $t = \text{const}$ происходят одновременно, а наблюдатель **M** должен придерживаться другого мнения: ему кажется, что одновременно происходят все события на каждом из «наклоненных» одновременных пространств! Геометрия Минковского сама по себе не содержит единственного понятия «одновременности»; но каждый наблюдатель, движущийся равномерно и прямолинейно, имеет свое собственное представление о том, что значит «одновременно».

Рассмотрим два события R и Q на рис. 5.21. С точки зрения наблюдателя **S** событие R происходит раньше события Q, так как R лежит в более раннем одновременном пространстве, чем Q. Но с точки зрения наблюдателя **M** все будет наоборот, и событие Q окажется в более раннем одновременном пространстве, чем R. Таким образом, для одного наблюдателя событие R происходит раньше события Q, а для другого наблюдателя — позже! (Так может случиться лишь потому, что события R и Q, как принято говорить, *пространственно разделены*, что означает следующее: каждое событие находится вне светового конуса другого события, в результате чего ни одна материальная частица или фотон не могут совершить путешествие от одного события к другому.) Даже при очень медленных относительных скоростях для точек, разделенных большими расстояниями, имеют место значительные различия в хронологической последовательности. Представим себе двух людей, медленно проходящих друг мимо друга на улице. События в туманности Андромеды (ближайшей большой галактики, находящейся на расстоянии 20 000 000 000 000 000 000 км от нашей собственной Галактики — Млечного Пути), одновременные по мнению этих двух прохожих, в тот момент, когда они поравняются друг с другом — могут отстоять по времени друг от друга на несколько суток (рис. 5.22). В то время как для одного из прохожих космический флот, отправленный с заданием уничтожить все живое на Земле, уже находится в полете, для

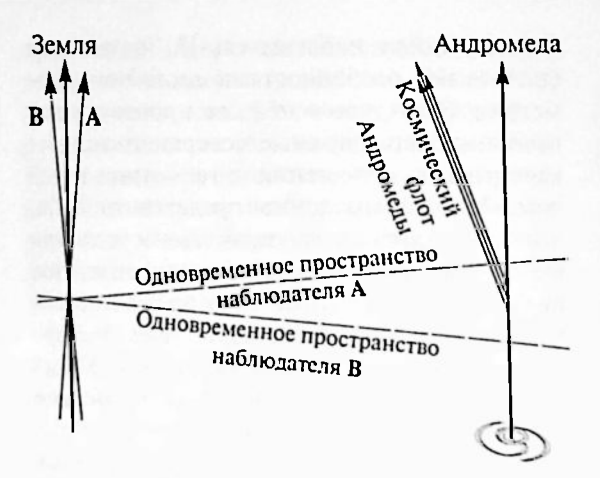


Рис. 5.22. Два наблюдателя **A** и **B** медленно проходят мимо друг друга. Их мнения относительно того, стартовал ли космический флот Андромеды в момент, когда они поравнялись, существенно отличаются

другого прохожего само решение относительно отправки космического флота в рейд еще не принято!

Общая теория относительности Эйнштейна

Напомним великую истину, открытую Галилеем: все тела под действием силы тяжести падают одинаково быстро. (Это было блестящей догадкой, едва ли подсказанной эмпирическими данными, поскольку из-за сопротивления воздуха перья и камни все же падают не *одновременно*! Галилей внезапно понял, что, если бы сопротивление воздуха можно было свести к нулю, то перья и камни *падали бы* на Землю одновременно.) Потребовалось три столетия, прежде чем глубокое значение этого открытия было по достоинству осознано и стало краеугольным камнем великой теории. Я имею в виду общую теорию относительности Эйнштейна — поразительное описание гравитации, для которого, как нам вскоре станет ясно, потребовалось введение понятия *искривленного пространства-времени*!

Какое отношение имеет интуитивное открытие Галилея к идее «кривизны пространства-времени»? Каким образом могло получиться, что эта концепция, столь явно отличная от схемы Ньютона, согласно которой частицы ускоряются под действием обычных гравитационных сил, оказалась

способной не только сравняться в точности описания с ньютоновской теорией, но и превзойти последнюю? И потом, насколько верным будет утверждение, что в открытии Галилея было нечто такое, что *не было* позднее включено в ньютоновскую теорию?

Позвольте мне начать с последнего вопроса потому, что ответить на него проще всего. Что, согласно теории Ньютона, управляет ускорением тела под действием гравитации? Во-первых, на тело действует гравитационная *сила*, которая, как гласит открытый Ньютоном закон всемирного тяготения, должна быть *пропорциональна массе тела*. Во-вторых, величина ускорения, испытываемая телом под действием *заданной* силы, по второму закону Ньютона, *обратно пропорциональна массе тела*. Удивительное открытие Галилея зависит от того факта, что «масса», входящая в открытый Ньютоном закон всемирного тяготения, есть, в действительности, та же «масса», которая входит во второй закон Ньютона. (Вместо «та же» можно было бы сказать «пропорциональна».) В результате ускорение тела под действием гравитации *не зависит* от его массы. В общей схеме Ньютона нет ничего такого, что указывало бы, что оба понятия массы одинаковы. Эту одинаковость Ньютон лишь *постулировал*. Действительно, электрические силы аналогичны гравитационным в том, что и те, и другие обратно пропорциональны квадрату расстояния, но электрические силы зависят от *электрического заряда*, который имеет совершенно другую природу, чем *масса* во втором законе Ньютона. «Интуитивное открытие Галилея» было бы неприменимо к электрическим силам: о телах (заряженных телах) брошенных в электрическом поле, нельзя сказать, что они «падают» с одинаковой скоростью!

На время просто *примем* интуитивное открытие Галилея относительно движения под действием *гравитации* и попытаемся выяснить, к каким следствиям оно приводит. Представим себе Галилея, бросающего с Пизанской наклонной башни два камня. Предположим, что с одним из камней жестко скреплена видеокамера, направленная на другой камень. Тогда на пленке окажется запечатленной следующая ситуация: камень парит в пространстве, как бы *не испытывая*

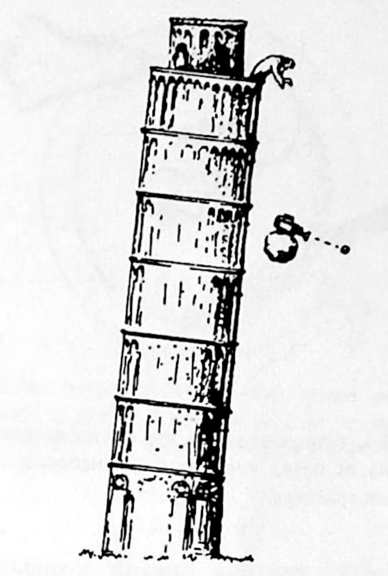


Рис. 5.23. Галилей бросает два камня (и видеокамеру) с Пизанской башни

действия гравитации (рис. 5.23)! И так происходит именно потому, что все тела под действием гравитации падают с одной и той же скоростью.

В описанной выше картине мы пренебрегаем сопротивлением воздуха. В наше время космические полеты открывают перед нами лучшую возможность проверки этих идей, так как в космическом пространстве нет воздуха. Кроме того, «падение» в космическом пространстве означает просто движение по определенной орбите под действием гравитации. Такое «падение» совсем не обязательно должно происходить по прямой вниз — к центру Земли. В нем вполне может быть и некоторая горизонтальная составляющая. Если эта горизонтальная составляющая достаточно велика, то тело может «падать» по круговой орбите вокруг Земли, не приближаясь к ее поверхности! Путешествие по свободной околоземной орбите под действием гравитации — весьма изощренный (и очень дорогой!) способ «падения». Как в описанной выше видеозаписи, астронавт, совершая «прогулку в открытом космосе», видит свой космический корабль парящим перед собой и как бы не испытывающим действия гравитации со стороны огромного шара Земли под ним! (См. рис. 5.24.) Таким образом, переходя в «ускоренную систему отсчета»

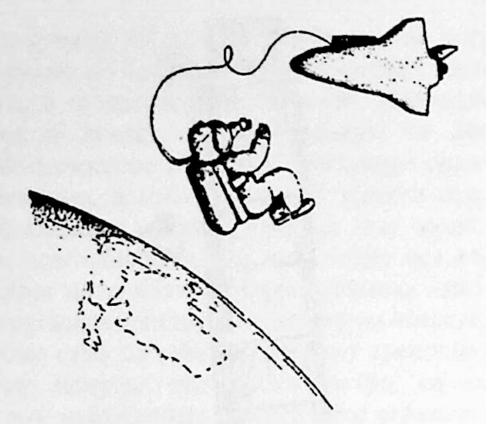


Рис. 5.24. Астронавт видит, что его космический корабль парит перед ним, как будто неподверженный действию гравитация

свободного падения, можно локально исключить действие гравитации.

Мы видим, что свободное падение позволяет *исключить* гравитацию потому, что эффект от действия гравитационного поля такой же, как от ускорения. Действительно, если вы находитесь в лифте, который движется с ускорением вверх, то вы просто ощущаете, что кажущееся гравитационное поле увеличивается, а если лифт движется с ускорением вниз, то вам кажется, что гравитационное поле убывает. Если бы трос, на котором подвешена кабина, оборвался, то (если пренебречь сопротивлением воздуха и эффектами трения) результирующее ускорение, направленное вниз (к центру Земли), полностью уничтожило бы действие гравитации, и люди, оказавшиеся в кабине лифта, стали бы свободно плавать в пространстве, подобно астронавту во время выхода в открытый космос, до тех пор, пока кабина не стукнулась бы о Землю! Даже в поезде или на борту самолета ускорения могут быть такими, что ощущения пассажира относительно величины и направления гравитации могут не совпадать с тем, где, как показывает обычный опыт, должны быть «верх» и «низ». Объясняется это тем, что действия ускорения и гравитации *схожи* настолько, что наши ощущения не способны отличить одни от других. Этот факт — то, что локальные проявления гравитации эквивалентны локальным проявлениям ускоренно движущейся системы отсчета, — и есть то, что Эйнштейн назвал *принципом эквивалентности.*

Приведенные выше соображения «локальны». Но если разрешается производить (не только локальные) измерения с достаточно высокой точностью, то в принципе можно установить *различие* между «истинным» гравитационным полем и чистым ускорением. На рис. 5.25 я изобразил в немного преувеличенном виде, как первоначально стационарная сферическая конфигурация частиц, свободно падающая под действием гравитации, начинает деформироваться под влиянием *неоднородности* (ньютоновского) гравитационного поля. Это поле неоднородно в двух отношениях. Во-первых, поскольку центр Земли расположен на некотором конечном расстоянии от падающего тела, частицы, расположенные ближе к поверхности Земли, движутся вниз с бóльшим ускорением, чем частицы, расположенные выше (напомним закон обратной пропорциональности квадрату расстояния Ньютона). Во-вторых, по той же причине существуют небольшие различия в направлении ускорения для частиц, занимающих различные положения на горизонтали. Из-за этой неоднородности сферическая форма начинает слегка деформироваться, превращаясь в «эллипсоид». Первоначальная сфера удлиняется в направлении к центру Земли (а также в противоположном направлении), так как те ее части, которые ближе к центру

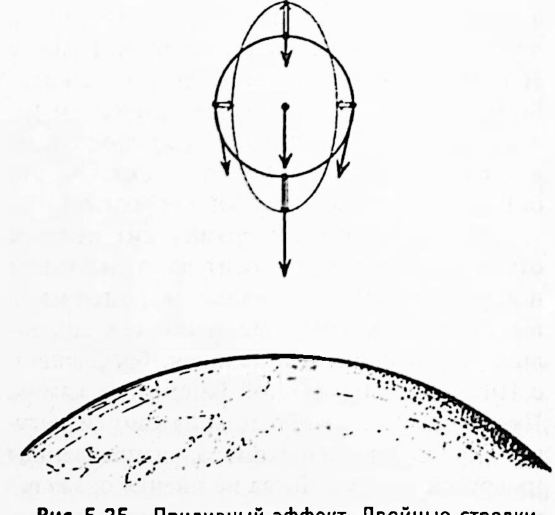


Рис. 5.25. Приливный эффект. Двойные стрелки указывают относительное ускорение (ВЕЙЛЬ)

Земли, движутся с чуть бо́льшим ускорением, чем те части, которые дальше от центра Земли, и сужается по горизонтали, так как ускорения ее частей, находящихся на концах горизонтального диаметра, слегка скошены «внутрь» — в направлении на центр Земли.

Это деформирующее действие известно как *приливный эффект* гравитации. Если мы заменим центр Земли Луной, а сферу из материальных частиц — поверхностью Земли, то получим в точности описание действия Луны, вызывающей приливы на Земле, причем «горбы» образуются по направлению к Луне и от Луны. Приливный эффект — общая особенность гравитационных полей, которая не может быть «исключена» с помощью свободного падения. Приливный эффект служит мерой неоднородности ньютоновского гравитационного поля. (Величина приливной деформации в действительности убывает обратно пропорционально *кубу*, а не квадрату расстояния от центра притяжения.)

Закон всемирного тяготения Ньютона, по которому сила обратно пропорциональна квадрату расстояния, допускает, как оказывается, простую интерпретацию в терминах приливного эффекта: *объем* эллипсоида, в который первоначально [18] деформируется сфера, *равен* объему исходной сферы — в предположении, что сфера окружает вакуум. Это свойство сохранения объема характерно для закона обратных квадратов; ни для каких других законов оно не выполняется. Предположим далее, что исходная сфера окружает не вакуум, а некоторое количество материи общей массой M. Тогда возникает дополнительная компонента ускорения, направленная внутрь сферы из-за гравитационного притяжения материи внутри сферы. Объем эллипсоида, в который первоначально деформируется наша сфера из материальных частиц, *сокращается* — на величину, *пропорциональную* M. С примером эффекта уменьшения объема эллипсоида мы бы столкнулись, если бы выбрали нашу сферу так, чтобы она окружала Землю на постоянной высоте (рис. 5.26). Тогда обычное ускорение, обусловленное земным притяжением и направленное вниз (т. е. внутрь Земли), будет той самой причиной, по которой происходит сокращение объема нашей сферы.

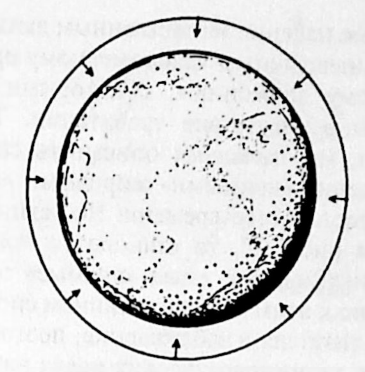


Рис. 5.26. Когда сфера окружает некое вещество (в данном случае — Землю), возникает результирующее ускорение, направленное внутрь (РИЧЧИ)

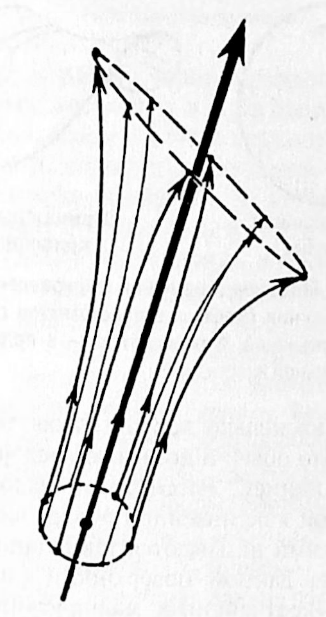


Рис. 5.27. Кривизна пространства-времени: приливный эффект, изображенный в пространстве-времени

В этом свойстве сжимания объема заключена оставшаяся часть закона всемирного тяготения Ньютона, а именно — что сила пропорциональна массе *притягивающего* тела.

Попробуем получить пространственно-временную картину такой ситуации. На рис. 5.27 я изобразил мировые линии частиц нашей сферической поверхности (представленной на рис. 5.25 в виде окружности), причем я использовал для описания ту систему отсчета, в которой центральная точка сферы кажется покоящейся («свободное падение»). Позиция общей теории относительности состоит в том, чтобы считать

свободное падение «естественным движением» — аналогичным «равномерному прямолинейному движению», с которыми имеют дело в отсутствие гравитации. Таким образом, мы *пытаемся* описывать свободное падение «прямыми» мировыми линиями в пространстве-времени! Но если взглянуть на рис. 5.27, то становится понятно, что использование *слова* «прямые» применительно к этим мировым линиям способно ввести читателя в заблуждение, поэтому мы будем в терминологических целях называть мировые линии свободно падающих частиц в пространстве-времени — *геодезическими*.

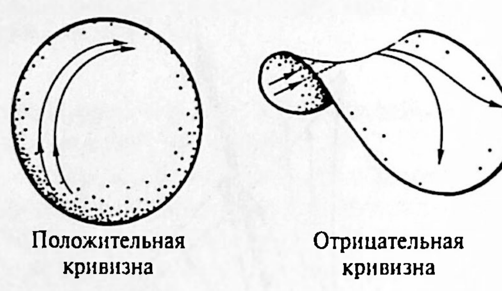


Рис. 5.28. Геодезические линии в искривленном пространстве: линии сходятся в пространстве с положительной кривизной, и расходятся — в пространстве с отрицательной кривизной

Но насколько хороша такая терминология? Что обычно понимают под «геодезической» линией? Рассмотрим аналогию для двумерной искривленной поверхности. Геодезическими называются такие кривые, которые на данной поверхности (локально) служат «кратчайшими маршрутами». Иначе говоря, если представить себе отрезок нити, натянутый на указанную поверхность (и не слишком длинный, чтобы он не мог соскользнуть), то нить расположится вдоль некоторой геодезической линии на поверхности. На рис. 5.28 я привел два примера поверхностей: первая (слева) — поверхность так называемой «положительной кривизны» (как поверхность сферы), вторая — поверхность «отрицательной кривизны» (седловидная поверхность). На поверхности положительной кривизны две соседние геодезические линии, выходящие из начальных точек параллельно друг другу, начинают впоследствии изгибаться *навстречу* друг другу; а на поверхности отрицательной кривизны они изгибаются *в стороны* друг *от* друга. Если мы представим себе, что мировые линии свободно падающих частиц в некотором смысле ведут себя как геодезические линии на поверхности, то окажется, что существует тесная аналогия между гравитационным приливным эффектом, о котором шла речь выше, и эффектами кривизны поверхности — причем как положительной кривизны, *так и* отрицательной. Взгляните на рис. 5.25, 5.27. Мы видим, что в нашем пространстве-времени геодезические линии начинают *расходиться* в одном направлении (когда они «выстраиваются» в сторону Земли) — как это происходит на поверхности *отрицательной* кривизны на рис. 5.28 — и *сближаться* в других направлениях (когда они смещаются горизонтально относительно Земли) — как на поверхности *положительной* кривизны на рис. 5.28. Таким образом, создается впечатление, что наше пространство-время, как и вышеупомянутые поверхности, тоже обладает «кривизной», только более сложной, поскольку из-за высокой размерности пространства-времени при различных перемещениях она может носить смешанный характер, не будучи ни чисто положительной, ни чисто отрицательной.

Отсюда следует, что понятие «кривизны» пространства-времени может быть использовано для описания действия гравитационных полей. Возможность использования такого описания в конечном счете следует из интуитивного открытия Галилея (принципа эквивалентности) и позволяет нам исключить гравитационную «силу» с помощью свободного падения. Действительно, ничто из сказанного мной до сих пор не выходит за рамки ньютонианской теории. Нарисованная только что картина дает просто *переформулировку* этой теории [19]. Но когда мы пытаемся скомбинировать новую картину с тем, что дает предложенное Минковским описание специальной теории относительности — геометрии пространства-времени, которая, как мы знаем, применяется в *отсутствие* гравитации — в игру вступает новая физика. Результат этой комбинации — *общая теория относительности* Эйнштейна.

Напомним, чему учил нас Минковский. Мы имеем (в отсутствие гравитации) про-

странство-время, наделенное особого рода мерой «расстояния» между точками: если мы имеем в пространстве-времени мировую линию, описывающую траекторию какой-нибудь частицы, то «расстояние» в смысле Минковского, измеряемое вдоль этой мировой линии, дает *время*, реально прожитое частицей. (В действительности, в предыдущем разделе мы рассматривали это «расстояние» только для тех мировых линий, которые состоят из прямолинейных отрезков — но приведенное выше утверждение справедливо и по отношению к искривленным мировым линиям, если «расстояние» измеряется вдоль кривой.) Геометрия Минковского считается точной, если нет гравитационного поля, т. е. если у пространства-времени нет кривизны. Но при наличии гравитации мы рассматриваем геометрию Минковского уже лишь как приближенную — аналогично тому, как плоская поверхность лишь приблизительно соответствует геометрии искривленной поверхности. Вообразим, что, изучая искривленную поверхность, мы берем микроскоп, дающий все большее увеличение — так, что геометрия искривленной поверхности кажется все больше растянутой. При этом поверхность будет нам казаться все более плоской. Поэтому мы говорим, что искривленная поверхность имеет локальное строение евклидовой плоскости [20]. Точно так же мы можем сказать, что при наличии гравитации пространство-время *локально* описывается геометрией Минковского (которая есть геометрия плоского пространства-времени), но мы допускаем некоторую «искривленность» на более крупных масштабах (рис. 5.29). В частности, как и в пространстве Минковского, любая точка пространства-времени является вершиной *светового конуса* — но в данном случае эти световые конусы расположены уже не одинаково. В главе 7 мы познакомимся с отдельными моделями пространства-времени, в которых явно видна эта неоднородность расположения световых конусов (см. рис. 7.13, 7.14 на с. 291, 293). Мировые линии материальных частиц всегда направлены *внутрь* световых конусов, а линии фотонов — *вдоль* световых конусов. Вдоль любой такой кривой мы можем ввести «расстояние» в смысле Минковского, которое

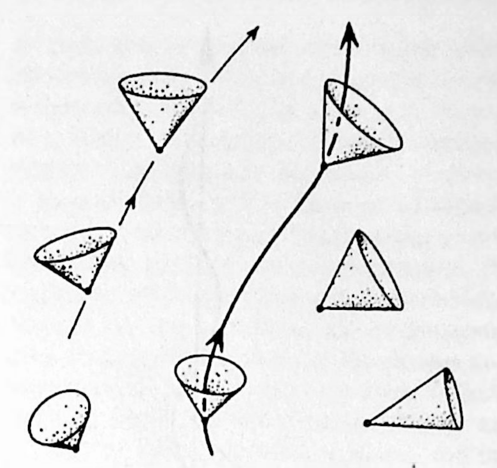


Рис. 5.29. Картина искривленного пространства-времени

служит мерой времени, прожитого частицами так же, как и в пространстве Минковского. Как и в случае искривленной поверхности, эта мера «расстояния» определяет *геометрию* поверхности, которая может отличаться от геометрии плоскости.

Геодезическим линиям в пространстве-времени теперь можно придать интерпретацию, аналогичную интерпретации геодезических линий на двумерных поверхностях, учитывая при этом различия между геометриями Минковского и Евклида. Таким образом, наши геодезические линии в пространстве-времени представляют собой не (локально) кратчайшие кривые, а наоборот — кривые, которые (локально) *максимизируют* «расстояние» (т. е. время) вдоль мировой линии. Мировые линии частиц, свободно перемещающиеся под действием гравитации, согласно этому правилу действительно *являются* геодезическими. В частности, небесные тела, движущиеся в гравитационном поле, хорошо описываются подобными геодезическими линиями. Кроме того, лучи света (мировые линии фотонов) в пустом пространстве так же служат геодезическими линиями, но на этот раз — *нулевой* «длины» [21]. В качестве примера я схематически нарисовал на рис. 5.30 мировые линии Земли и Солнца. Движение Земли вокруг Солнца описывается «штопорообразной» линией, навивающейся вокруг мировой линии Солнца. Там же я изобразил фотон, приходящий на Землю от далекой звезды. Его мировая линия кажется слегка

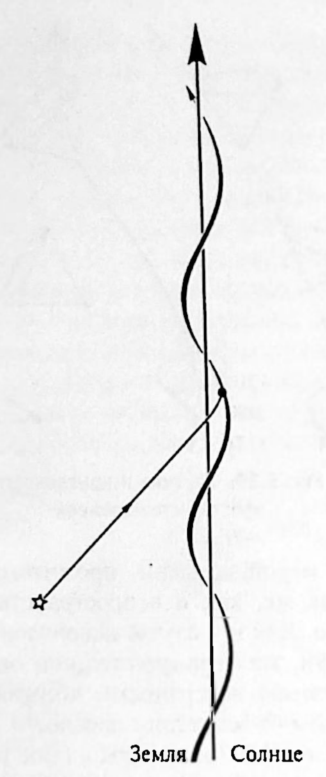


Рис. 5.30. Мировые линии Земли и Солнца. Световой луч от далекой звезды отклоняется Солнцем

«изогнутой» вследствие того, что свет (по теории Эйнштейна) на самом деле отклоняется гравитационным полем Солнца.

Нам необходимо еще выяснить, каким образом ньютоновский закон обратных квадратов может быть включен (после надлежащей модификации) в общую теорию относительности Эйнштейна. Обратимся еще раз к нашей сфере из материальных частиц, падающей в гравитационном поле. Напомним, что если внутри сферы заключен только вакуум, то, согласно теории Ньютона, объем сферы первоначально не изменяется; но если внутри сферы находится материя общей массой M, то происходит сокращение объема, пропорциональное M. В теории Эйнштейна (для малой сферы) правила в точности такие же, за исключением того, что не все изменение объема определяется массой M; существует (обычно очень малый) вклад от *давления*, возникающем в окруженном сферой материале.

Полное математическое выражение для кривизны четырехмерного пространства-времени (которая должна описывать приливные эффекты для частиц, движущихся в любой данной точке по всевозможным направлениям) дается так называемым *тензором кривизны Римана*. Это несколько сложный объект; для его описания необходимо в каждой точке указать двадцать действительных чисел. Эти двадцать чисел называются его *компонентами*. Различные компоненты соответствуют различным кривизнам в различных направлениях пространства-времени. Тензор кривизны Римана обычно записывают в виде R_{ijkl}, но так как мне не хочется объяснять здесь, что означают эти субиндексы (и, конечно, что такое тензор), то я запишу его просто как:

$$\text{РИМАН}.$$

Существует способ, позволяющий разбить этот тензор на две части, называемые, соответственно, тензором ВЕЙЛЯ и тензором РИЧЧИ (каждый — с десятью компонентами). Условно я запишу это разбиение так:

$$\text{РИМАН} = \text{ВЕЙЛЬ} + \text{РИЧЧИ}.$$

(Подробная запись тензоров Вейля и Риччи для наших целей сейчас совершенно не нужна.) Тензор Вейля ВЕЙЛЬ служит мерой *приливной деформации* нашей сферы из свободно падающих частиц (т. е. изменения начальной формы, а не размеров); тогда как тензор Риччи РИЧЧИ служит мерой изменения первоначального объема [22]. Напомним, что ньютоновская теория гравитации требует, чтобы *масса*, содержащаяся внутри нашей падающей сферы, была пропорциональна этому изменению первоначального объема. Это означает, что, грубо говоря, плотность *массы* материи — или, что эквивалентно, плотность *энергии* (так как $E = mc^2$) — следует *приравнять* тензору Риччи.

По существу, это именно то, что утверждают уравнения поля общей теории относительности, а именно — *полевые уравнения Эйнштейна* [23]. Правда, здесь имеются некоторые технические тонкости, в которые нам сейчас, впрочем, лучше не вдаваться. Достаточно сказать, что существует объект, называемый тензором *энергии-импульса*, который объединяет всю существенную информацию об энергии, давлении и импульсе материи и электромагнитных полей. Я буду называть этот тензор ЭНЕРГИЕЙ. Тогда уравнения

Эйнштейна весьма схематично можно представить в следующем виде,

РИЧЧИ = ЭНЕРГИЯ.

(Именно наличие «давления» в тензоре ЭНЕРГИЯ вместе с некоторыми требованиями непротиворечивости уравнений в целом приводят с необходимостью к учету давления в описанном выше эффекте сокращения объема.)

Кажется, что вышеприведенное соотношение ничего не говорит о тензоре Вейля. Тем не менее, оно отражает одно важное свойство. Приливный эффект, производимый в пустом пространстве, обусловлен ВЕЙЛЕМ. Действительно, из приведенных выше уравнений Эйнштейна следует, что существуют *дифференциальные* уравнения, связывающие ВЕЙЛЯ с ЭНЕРГИЕЙ — практически как во встречавшихся нам ранее уравнениях Максвелла [24]. Действительно, точка зрения, согласно которой ВЕЙЛЯ надлежит рассматривать как своего рода гравитационный аналог электромагнитного поля (в действительности, тензора — тензора Максвелла), описываемого парой (**E**, **B**), оказывается весьма плодотворной. В этом случае ВЕЙЛЬ служит своего рода мерой *гравитационного* поля. «Источником» для ВЕЙЛЯ является ЭНЕРГИЯ — подобно тому, как источником для электромагнитного поля (**E**, **B**) является $(\rho, \mathbf{j})$ — набор из зарядов и токов в теории Максвелла. Эта точка зрения будет полезна нам в главе 7.

Может показаться весьма удивительным, что при столь существенных различиях в формулировке и основополагающих идеях, оказывается довольно трудно найти наблюдаемые различия между теориями Эйнштейна и теорией, выдвинутой Ньютоном двумя с половиной столетиями раньше. Но если рассматриваемые скорости малы по сравнению со скоростью света c, а гравитационные поля не слишком сильны (так, что скорости убегания гораздо меньше c, см. главу 7, с. 158), то теория Эйнштейна по существу дает те же результаты, что и теория Ньютона. Но в тех ситуациях, когда предсказания этих двух теорий расходятся, прогнозы теории Эйнштейна оказываются точнее. К настоящему времени был проведен целый ряд весьма впечатляющих экспериментальных проверок, которые позволяют считать новую теорию Эйнштейна вполне обоснованной. Часы, согласно Эйнштейну, в гравитационном поле идут чуть медленнее. Ныне этот эффект измерен непосредственно несколькими способами. Световые и радиосигналы действительно изгибаются вблизи Солнца и слегка запаздывают для наблюдателя, движущегося им навстречу. Эти эффекты, предсказанные изначально общей теорией относительности, на сегодняшний день подтверждены опытом. Движение космических зондов и планет требуют небольших поправок к ньютоновским орбитам, как это следует из теории Эйнштейна — эти поправки сегодня также проверены опытным путем. (В частности, аномалия в движении планеты Меркурия, известная как «смещение перигелия», беспокоившая астрономов с 1859 года, была объяснена Эйнштейном в 1915 году.) Возможно, наиболее впечатляющим из всего следует считать серию наблюдений над системой, называемой *двойным пульсаром*, которая состоит из двух небольших массивных звезд (возможно, двух «нейтронных звезд», см. с. 290). Эта серия наблюдений очень хорошо согласуется с теорией Эйнштейна и служит прямой проверкой эффекта, полностью отсутствующего в теории Ньютона, — испускания *гравитационных волн*. (Гравитационная волна представляет собой аналог электромагнитной волны и распространяется со скоростью света c.) Не существует проверенных наблюдений, которые противоречили бы общей теории относительности Эйнштейна. При всей своей странности (на первый взгляд), теория Эйнштейна работает и по сей день!

Релятивистская причинность и детерминизм

Напомним, что в теории относительности материальные тела не могут двигаться быстрее света — откуда, в частности, следует, что их мировые линии всегда должны лежать внутри световых конусов (см. рис. 5.29). (В общей теории относительности ситуацию следует формулировать именно в таком локальном виде. Световые конусы расположены неодинаково, поэтому не имело бы особого смысла говорить, превосходит ли

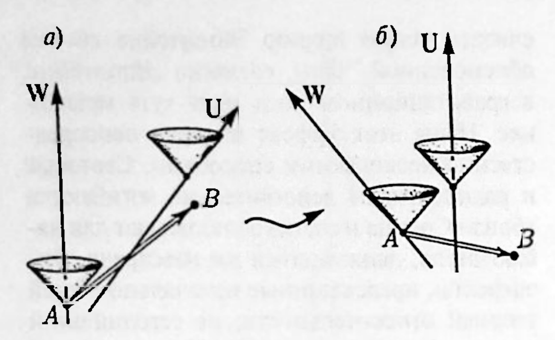


Рис. 5.31. Сигнал, который распространяется для наблюдателя W быстрее света, для наблюдается U распространяется назад по времени. Ситуация справа (*б*) представляет собой ту же ситуацию, что и слева (*а*), только перерисованную с точки зрения наблюдателя U. (Эту перерисовку можно рассматривать как движение Пуанкаре. Сравните с рис. 5.21 — но здесь преобразование от (*а*) к (*б*) следует понимать в *активном*, а не в пассивном смысле.)

скорость *очень далекой* частицы скорость света *здесь.*) Мировые линии фотонов проходят *по поверхности* световых конусов, но мировая линия ни одной частицы не должна лежать *вне* световых конусов. В действительности, должно выполняться более общее утверждение, а именно: ни одному сигналу не разрешается распространяться вне светового конуса.

Чтобы понять, почему должно быть именно так, рассмотрим снова картину пространства Минковского (рис. 5.31). Предположим, что сконструировано некоторое устройство, способное посылать сигнал со скоростью немного больше скорости света. Пользуясь этим устройством, наблюдатель W посылает сигнал из точки A на своей мировой линии к далекой точке B, расположенной непосредственно под световым конусом события A. На рис. 5.31 *а* эта ситуация изображена с точки зрения наблюдателя W, но на рис. 5.31 *б* картина нарисована уже по-другому, с точки зрения второго наблюдателя U, который быстро движется от W (из точки, например, между A и B) — и наблюдателю U событие B кажется происходящим *раньше* события A! (Такая «перерисовка» есть не что иное, как движение Пуанкаре, как описано выше, см. с. 187.) С точки зрения наблюдателя W одновременные пространства наблюдателя U представляются «наклоненными». Поэтому событие B кажется наблюдателю U происходящим раньше события A. Таким образом, для U сигнал, испущенный наблюдателем W, будет распространяться назад во времени!

Здесь пока еще нет явного противоречия. Но, учитывая симметричность картины с точки зрения наблюдателя U (в силу принципа специальной относительности), *третий* наблюдатель V, движущийся от наблюдателя U в сторону, противоположную той, в которую движется наблюдатель W, и оснащенный таким же, как и у наблюдателя W, устройством, мог бы в свою очередь послать сигнал, распространяющийся быстрее света с его (наблюдателя V) точки зрения, в направлении, противоположном направлению сигнала, испущенного наблюдателем W. Наблюдателю U при этом будет казаться, что сигнал, испущенный наблюдателем V, тоже движется назад во времени — но в противоположном (пространственном) направлении. Действительно, наблюдатель V мог бы послать второй сигнал к наблюдателю W в момент (B) получения исходного сигнала, пришедшего от наблюдателя W. Этот сигнал достигает наблюдателя W в тот момент, когда происходит событие C, которое (по оценке наблюдателя U) предшествует испусканию исходного сигнала (событию A) (рис. 5.32). Но еще хуже то, что событие C действительно происходит раньше события A (испускания исходного сигнала) на *собственной мировой* линии наблюдателя W, поэтому W действительно *воспринимает* событие C как происходящее

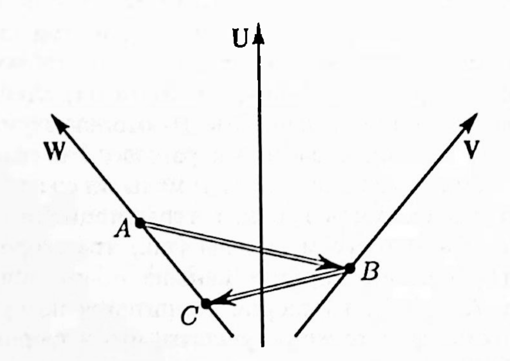


Рис. 5.32. Если у наблюдателя V имеется сверхсветовое сигнальное устройство, тождественное устройству, имеющемуся у W, но посылающее сигналы в противоположном направлении, то наблюдатель W может им воспользоваться для того, чтобы отправить послание в свое собственное прошлое!

до того, как он испускает сигнал (события A)! Сигнал, отправляемый наблюдателем **V** обратно наблюдателю **W**, мог бы, по предварительной договоренности с **W**, просто повторять сигнал, полученный наблюдателем **W** в точке B. Таким образом, **W** получает в более ранний момент времени на своей мировой линии тот же самый сигнал, который он сам собирается послать позднее! Разнося двух наблюдателей достаточно далеко друг от друга, можно устроить все так, что ответный сигнал будет опережать исходный на сколь угодно большое время. Возможно, наблюдатель **W** своим исходным сигналом сообщал о том, что он сломал ногу. Тогда ответный сигнал он мог бы получить задолго *до* того, как с ним произошло это печальное происшествие, и тогда (предположительно) он мог бы предпринять необходимые меры предосторожности и избежать несчастного случая!

Таким образом, распространение сигналов со сверхсветовыми скоростями вместе с эйнштейновским принципом относительности приводит к вопиющему противоречию с нашим нормальным пониманием «свободы воли». В действительности, ситуация еще более серьезна, чем до сих пор представлялось. Ибо мы могли бы сделать «наблюдателя **W**» всего лишь механическим устройством, запрограммированным так, чтобы посылать в ответ тот же сигнал, который был им получен (т.е. отвечать на «**НЕТ**» — «**НЕТ**» и на «**ДА**» — «**ДА**»). Это приводит к такому же принципиальному противоречию, как то, с которым нам уже приходилось сталкиваться прежде [25]. Причем кажется, что на этот раз оно не зависит от наличия у наблюдателя **W** «свободы воли». Это свидетельствует о том, что на устройство, способное испускать сверхсветовые сигналы, не стоит «делать ставку» как на физически возможное. В дальнейшем это обстоятельство еще приведет нас с вами к удивительным выводам (глава 6, с. 252).

Исходя из вышесказанного, давайте примем, что сигналы *любого* рода — а не только переносимые обычными физическими частицами — должны быть ограничены световыми конусами. Действительно, то, о чем мы только что говорили, опирается на идеи *специальной* теории относительности — но и в общей теории относительности правила СТО (локально) остаются в силе. Именно локальная выполнимость положений специальной теории относительности позволяет утверждать, что все сигналы остаются в пределах световых конусов, поэтому то же самое должно выполнятся и в общей теории относительности. Далее мы посмотрим, как это отражается на вопросах *детерминизма* в рамках этих теорий. Напомним, что в ньютоновской (или гамильтоновой и т.д.) схеме «детерминизм» — это возможность однозначного определения поведения системы в любой момент времени при условии, что заданы *начальные условия*. Если мы будем смотреть на ньютоновскую теорию с точки зрения пространства-времени, то «конкретное время», когда мы задаем эти начальные условия, будет представлено некоторым трехмерным «слоем» в четырехмерном пространстве-времени (т. е. будет всем пространством в этот момент времени). В теории относительности не существует одного глобального понятия «времени», которое можно было бы выделить для этой цели. Обычный подход предполагает гибкое отношение к этому вопросу. Годится любое «время». В специальной теории относительности вместо упоминавшегося выше «слоя» можно взять одновременное пространство какого-нибудь наблюдателя и задать на нем начальные данные. Но в общей теории относительности понятие «одновременного пространства» достаточно размыто. Вместо него можно воспользоваться более общим понятием *пространственно-подобной поверхности* [26]. Такая поверхность изображена на рис. 5.33; она характеризуется тем, что в каждой из своих точек она лежит целиком вне светового конуса — так,

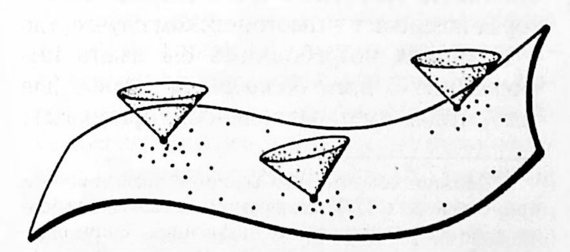


Рис. 5.33. Пространственно-подобная поверхность для задания начальных условий в общей теории относительности



Рис. 5.34. В специальной теории относительности то, что происходит в точке P, зависит только от данных, заданных в *конечной* области одновременного пространства. Так происходит потому, что никакое воздействие не может достичь точки P быстрее света

что *локально* она напоминает одновременное пространство.

Детерминизм в СТО можно сформулировать так: начальные данные на любом заданном одновременном пространстве S определяют поведение системы во всем пространстве-времени. (В частности, это верно для теории Максвелла, которая действительно является «специально релятивистской» теорией.) Однако можно высказать и более сильное утверждение. Если мы хотим знать, что произойдет в некоторой точке P, лежащей где-то в будущем по отношению к пространству S, то для этого нам необходимы начальные данные не на всем S, а только в некоторой ограниченной (конечной) области пространства S — потому, что «информация» не может распространяться быстрее света, так что любые точки пространства S, лежащие слишком далеко для того, чтобы световые сигналы из них могли достигать P, не оказывают на P никакого влияния (рис. 5.34)[14]. Это гораздо более удовлетворительный результат по сравнению с той ситуацией, которая возникает в ньютоновском случае, где в принципе потребовалось бы иметь информацию о всем *бесконечном* «слое», для того, чтобы иметь возможность предсказать ближайшее будущее хотя бы для одной точки. На скорость, с которой может распространяться ньютоновская информация, не существует никаких ограничений, и действие ньютоновских сил поэтому распространяется *мгновенно*.

«Детерминизм» в *общей* теории относительности — вопрос гораздо более сложный, чем в СТО, и я ограничусь здесь лишь несколькими замечаниями. Прежде всего, для задания начальных условий нам необходимо воспользоваться *пространственноподобной поверхностью* S (а не просто одновременной поверхностью). Тогда оказывается, что уравнение Эйнштейна задают *локально* детерминистское поведение гравитационного поля в предположении (как обычно), что поля материи, дающие вклад в тензор ЭНЕРГИЯ, ведут себя детерминистским образом. Однако здесь возникают значительные осложнения. Сама геометрия пространства-времени (включая ее «причинную» структуру — расположение световых конусов) теперь становится частью того, что требуется определить. Априори расположение световых конусов нам не известно, так что мы не можем сказать, какие части поверхности S необходимы для однозначного определения поведения системы в некотором будущем событии P. Но могут сложиться такие экстремальные ситуации, когда *всех* точек поверхности S для этого окажется недостаточно, и, соответственно, глобальный детерминизм будет утрачен! (Здесь затрагиваются непростые вопросы, имеющие отношение к одной важной нерешенной пока проблеме в общей теории относительности, которая известна под названием «космической цензуры» и связана с образованием *черных дыр* (Типлер и др. [1980]); см. главу 7, с. 291, а также примечание на с. 293 и с. 298.) Маловероятно, чтобы любое подобное «крушение детерминизма», обусловленное «экстремальными» гравитационными полями, имело непосредственное отношение к тому, что происходит на «человеческих» масштабах — но тем не менее это недвусмысленно указывает на отсутствие ясности в вопросе о детерминизме в рамках общей теории относительности.

---

[14] Можно заметить, что волновое уравнение (см. примечание на с. 177), как и уравнение Максвелла, также является релятивистским уравнением. Таким образом, «феномен невычислимости» Пур-Эля—Ричардса, рассмотренный нами ранее, тоже зависит только от начальных данных в ограниченных областях пространства S.

Вычислимость в классической физике: где мы находимся?

На протяжении всей этой главы я старался не упускать из виду проблему *вычислимости* и, проводя различие между вычислимостью и детерминизмом, стремился показать, что первая может иметь не меньшее значение, коль скоро речь заходит о «свободе воли» и умственной деятельности. Но само понятие детерминизма в рамках классической теории оказалось не настолько четко определенным, как принято было думать. Мы видели, что при изучении классического уравнения Лоренца для движения заряженной частицы возникает целый ряд тревожных вопросов. (Вспомним «убегающие решения» Дирака.) Потом было показано, что и в общей теории относительности с детерминизмом сопряжены определенные трудности. Когда в таких теориях нет детерминизма — в них заведомо нет и вычислимости. Тем не менее ни в одном из названных случаев не создается впечатление, что отказ от детерминизма может существенным образом повлиять на нашу философию. В подобных явлениях еще «нет места» для нашей свободы воли: во-первых, потому, что классическое уравнение Лоренца для точечной частицы (в том виде, как его решил Дирак) нельзя считать пригодным с физической точки зрения для использования на том уровне, где возникают эти проблемы; и, во-вторых, потому, что масштабы, на которых классическая общая теория относительности приводит к такого рода проблемам (черные дыры и т. д.), в принципе не сравнимы с масштабами нашего собственного головного мозга.

Спрашивается: что мы сейчас знаем о *вычислимости* в классической теории? Разумно предположить, что в общей теории относительности мы сталкиваемся с теми же проблемами, что и в СТО — если не считать тех различий в вопросах причинности и детерминизма, о которых было только что сказано. Там, где будущее поведение физической системы определяется начальными данными, оно в то же время должно (из соображений, изложенных при рассмотрении ньютоновской теории) быть *вычислимо* на основе тех же начальных данных [27] (не считая «бесполезного» типа невычислимости, с которым столкнулись Пур-Эль и Ричардс в случае волнового уравнения, о чем уже говорилось выше; эта ситуация не реализуется при *гладко* изменяющихся данных). Действительно, трудно представить, каким образом в *любой* из рассмотренных мной до сих пор физических теорий могут возникнуть какие-либо существенные «невычислимые» элементы. Можно заведомо предсказать, что «хаотическое» поведение является типичным для большинства из этих теорий, где весьма малые изменения начальных данных способны вызвать громадные расхождения в последующем поведении. (Именно так, насколько можно судить, обстоит дело в общей теории относительности; см. Мизнер [1969], Белинский и др. [1970].) Но, как я уже упоминал выше, довольно трудно понять, каким образом *этот* тип невычислимости (т. е. непредсказуемости) может быть «использован» в устройстве, с помощью которого мы могли бы попытаться «подчинить» себе возможные невычислимые элементы в физических законах. Если «разум» способен каким-то образом использовать невычислимые элементы, то последние должны, видимо, лежать вне классической физики. Нам придется еще раз вернуться к этому вопросу позднее — после того, как мы в общих чертах познакомимся с квантовой теорией.

Масса, материя и реальность

Произведем небольшую «ревизию» той картины мира, которую дала нам классическая физика. Во-первых, там существует пространство-время, выполняющее важнейшую функцию арены, на которой разыгрываются всевозможные физические процессы. Во-вторых, имеются *физические объекты*, задействованные в этих процессах, но ограниченные точными математическими законами. Физические объекты, о которых идет речь, бывают двух типов: *частицы* (*корпускулы*) и *поля*. Об истинной природе и отличительных особенностях частиц сказано немного, за исключением того, что у каждой частицы

«Длина» этой стрелки в смысле Минковского — это важная величина, известная как *масса покоя*. Она описывает массу объекта в системе отсчета наблюдателя, покоящегося относительно этого объекта. Можно было бы рассматривать такую величину в качестве хорошей меры «количества материи», входящей в состав указанного объекта. Но подобная величина не аддитивна: если систему разделить на две, то исходная масса покоя не равна сумме масс покоя возникших в результате деления частей. Напомним рассмотренный выше распад π°-мезона. π°-мезон имеет положительную массу покоя, тогда как масса покоя каждого из возникших в результате распада фотонов равна нулю. Но свойство аддитивности выполняется для всей стрелки (четырехвектора), по отношению к которой мы должны выполнять «сложение» *векторного* типа, как показано на рис. 5.6. Именно *вся стрелка* служит мерой «количества материи»!

Обратимся теперь к электромагнитному полю Максвелла. Мы уже отмечали, что оно переносит энергию. Значит, по соотношению $E = mc^2$ электромагнитное поле должно тоже иметь массу. Таким образом, и поле Максвелла представляет собой материю! И с этим утверждением теперь придется согласится, коль скоро поле Максвелла тесно связано с силами, удерживающими частицы вместе. Электромагнитные поля внутри любого тела должны вносить существенный вклад [28] в его массу.

А как обстоит дело с гравитационным полем Эйнштейна? Во многих отношениях оно напоминает поле Максвелла. Подобно тому, как в теории Максвелла заряженные тела, двигаясь, могут испускать *электромагнитные* волны, массивные движущиеся тела тоже могут (согласно теории Эйнштейна) порождать *гравитационные волны* (см. с. 195), которые, как и электромагнитные волны, распространяются со скоростью света, перенося при этом энергию. Однако эта энергия не поддается измерению стандартным способом, т. е. не может быть определена тензором ЭНЕРГИЯ, о котором говорилось выше. Для (чисто) гравитационной волны этот тензор всюду равен *нулю*! Можно было бы принять точку зрения, согласно которой *кривизна* пространства-времени (не полностью задаваемая тензором ВЕЙЛЬ) может каким-то образом представлять «количество материи», заключенной в гравитационных волнах. Но оказывается, что гравитационная энергия *нелокальна*: изучая кривизну пространства-времени только в ограниченных областях, невозможно определить, какова мера гравитационной энергии. Энергия, а следовательно, и масса гравитационного поля ведут себя подобно скользкому угрю, так что их невозможно «привязать» в каком-нибудь четко определенному месту. Тем не менее, к гравитационной энергии следует относиться со всей серьезностью. Она заведомо присутствует, и ее необходимо учитывать для того, чтобы сохранить смысл понятия массы. Существует хорошая (и положительная) мера массы (Бонди [1960] и Сакс [1962]), которая применима к гравитационным волнам — но нелокальность такова, что, как оказывается, эта мера может иногда становиться *ненулевой* в *плоских* областях пространства-времени, расположенных между двумя всплесками излучения (совсем как «глаз» урагана), где пространство-время на самом деле полностью *лишено* кривизны (см. Пенроуз, Риндлер [1986]) (и где, следовательно, оба тензора — ВЕЙЛЬ и РИЧЧИ — равны нулю)! В таких случаях мы, по-видимому, вынуждены придти к заключению, что если эта масса-энергия вообще должна быть локализована, то она с необходимостью должна быть сосредоточена в этом *плоском пустом пространстве* — области, совершенно свободной от материи или полей любого рода. При таких любопытных обстоятельствах наше «количество материи» либо локализовано *там*, в самых пустых областях пустого пространства — либо ее вообще нигде нет!

Такое заключение кажется чистейшим парадоксом. Но мы знаем, что этот вывод непосредственно вытекает из тех сведений о природе «реальной» материи нашего мира, которые дают наши лучшие классические теории (а это действительно превосходные теории!). Согласно классической теории — не говоря уже о квантовой, к изучению которой мы скоро приступим — материальная реальность оказывается субстанцией гораздо более расплывчатой, чем казалось прежде. Задача ее количественного измерения —

и даже само ее существование — связана с необходимостью учета чрезвычайно тонких моментов и не может быть выполнена только локально! Если такая нелокальность кажется вам загадочной — приготовьтесь к еще более сильным потрясениям!

Примечания

1. Поразительно, что *все* установленные отклонения от ньютоновской картины связаны фундаментальным образом с поведением света. Во-первых, это существование бестелесных полей, переносящих энергию, которые описываются электромагнитной теорией Максвелла. Во-вторых, как мы увидим, скорость света играет решающую роль в специальной теории относительности Эйнштейна. В-третьих, незначительные отклонения от ньютоновской теории гравитации, о которых нам говорит эйнштейновская *общая* теория относительности, становятся существенными только при скоростях, сравнимых со скоростью света. (Отклонение света вблизи Солнца, движение Меркурия, скорости убегания, сравнимые со скоростью убегания света для черных дыр, и т. д.) В-четвертых, дуализм волна-частица в квантовой теории впервые был обнаружен в поведении света. Наконец, нельзя не упомянуть о квантовой электродинамике — квантовой полевой теории света и заряженных частиц. Можно легко представить себе, что сам Ньютон с готовностью согласился бы признать, что фундаментальные проблемы его картины мира кроются в загадочном поведении света (см. Ньютон [1730], а также Пенроуз [1987a]).

2. Существует величественная область давно устоявшегося физического знания, а именно — *термодинамика* Карно, Максвелла, Кельвина, Больцмана и других ученых, которую я не собираюсь здесь классифицировать. Мое решение может показаться некоторым из читателей странным, но я поступаю так умышленно. По причинам, которые, возможно, станут яснее в главе 7, я не испытывал ни малейшего желания заносить термодинамику в том виде, какой она имеет в настоящее время, в категорию истинно ПРЕВОСХОДНЫХ теорий. Но многие физики, вероятно, сочли бы *кощунством*, осмелься я назвать это великолепное собрание таких красивых фундаментальных идей унизительным термином ПОЛЕЗНЫЕ! Я считаю, что термодинамика в ее обычном понимании — как дисциплины, оперирующей только *средними* величинами и ничего не говорящая о поведении отдельных компонентов системы; науки, отчасти сочетающей в себе следствия из других теорий, — не является в полном смысле физической теорией (конечно, с моей точки зрения, которую я распространяю и на математический базис *статистической механики*). Пользуясь случаем, я все же приношу читателям извинения за то, что оставил в стороне эту проблему и предпочел оставить вопрос классификации в стороне. Далее в главе 7 я высказываю уверенность в том, что между термодинамикой и теми идеями, которые относятся к модели Большого взрыва и которые я ранее охарактеризовал как ПОЛЕЗНЫЕ, существует тесная взаимосвязь. Убежден, что в результате правильного объединения этих двух теорий (пока еще, увы, не реализованного), должна возникнуть новая теория, которую с полным основанием можно будет отнести к категории ПРЕВОСХОДНЫХ. К этому вопросу нам еще придется вернуться в дальнейшем.

3. Мои коллеги спросили меня, в какую категорию я поместил бы «теорию твисторов» — глубоко разработанный круг идей и процедур, с которыми я был связан на протяжении долгих лет. Поскольку она представляет собой *альтернативную* теорию окружающего мира, ее нельзя охарактеризовать иначе, как ПРОБНУЮ; но, по большому счету, она является просто способом математической записи давно созданных физических теорий.

4. Но, по-видимому, Галилей часто использовал в своих наблюдениях водяные часы для измерения времени (см. Барба [1989]).

5. Эту модель связывают с именем Ньютона, но, как и в случае с «ньютоновской» механикой в целом, это — всего лишь удобный *ярлык*. Собственные взгляды Ньютона на *истинную* природу физического мира, по-видимому, отличались куда меньшим догматизмом и куда большей гибкостью. (Наиболее ярым сторонником «ньютоновской» модели, как представляется, был Р. Г. Бошкович (1711–1787).)

6. Рафаил Соркин разъяснил мне, что в некотором смысле эволюция этой конкретной игрушечной модели может быть сделана «вычислимой» в целом таким же способом, как (скажем) ньютоновские системы. Рассмотрим последовательность вычислений C_1, C_2, C_3, ..., которые позволят нам рассчитывать поведение нашей системы в (неограниченном) будущем со все возрастающей точностью (см. с. 166). В данном случае мы можем предположить, что C_N определяется при помощи машины Тьюринга, которая выполняет действие $T_u(m)$ в течение N шагов, и *положить* $T_u(m) = \square$, если она не остановилась на N-ом шаге. Однако, было бы нетрудно модифицировать нашу игрушечную модель таким образом, чтобы провалить подобные «вычисления» — для этого достаточно рассмотреть эволюцию, где выражение $T_u(m) = \square$ заменено на дважды квантифицированные утверждения вроде «$T(q)$ останавливается при всех q». (Нерешенная задача, связанная с наличием бесконечного множества пар простых чисел, отличающихся на «2», может служить примером такого утверждения.)

7. В главе 4 (примечание 10, с. 146) высказывалось предположение о том, что теория Блюма—Шуба—Смэйла [1989], вероятно, даст возможность решить некоторые из этих вопросов в математически более приемлемом виде.

8. Уравнения, написанные Гамильтоном, — хотя, возможно, не вполне отражавшие его собственную точку зрения — были известны великому итало-французскому математику Жозефу Л. Лагранжу (1736–1813) еще за 24 года до Гамильтона. Не менее важным достижением стала примерно в то же время формулировка механики в форме *уравнений Эйлера—Лагранжа*, согласно которым законы Ньютона можно рассматривать как производные одного основополагающего принципа — *принципа стационарного действия* (П. Л. М. де Мопертюи.) Обладая огромным теоретическим значением, уравнения Эйлера—Лагранжа имеют к тому же и немалую практическую ценность как мощный инструмент для вычислений.

9. В действительности, ситуация еще более «осложняется» в результате того, что лиувиллевский объем в фазовом пространстве — всего лишь один из целого семейства «объемов» различного числа измерений (называемых инвариантами Пуанкаре), которые остаются постоянными в ходе эволюции системы, описываемой уравнениями Гамильтона. Однако я был немного несправедлив в оценке всеобщности моих утверждений. Можно представить себе систему, в которой физические степени свободы (дающие вклад в какой-то из объемов фазового пространства) могут быть «заброшены» за пределы области наших интересы (например, они могут относиться к излучению, уходящему на бесконечность), так что объем той части фазового пространства, которую мы непосредственно изучаем, мог бы, на самом деле, уменьшиться.

10. Этот второй факт следует считать исключительной удачей для науки, ибо без него динамическое поведение больших тел могло бы остаться непостижимым и никак не указывало бы на конкретный вид тех законов, которые управляют поведением отдельных частиц. Как мне кажется, Ньютон столь упорно настаивал на своем третьем законе в том числе и потому, что без третьего закона динамическое поведение было бы просто невозможно перенести с микроскопического уровня на макроскопический.
Наряду с этим, не менее важное значение для развития естествознания имело еще одно «чудесное» совпадение, касающееся закона обратных квадратов: оказалось, что этот закон — единственный из всех степенных законов (описывающих убывающие с расстоянием силы) для которого орбиты движения вокруг центрального тела в общем случае имеют простую геометрическую форму. Что делал бы Кеплер, если бы сила всемирного тяготения была бы обратно пропорциональна не квадрату, а кубу расстояния?

11. Я выбрал единицы для различных полей так, чтобы они находились в хорошем согласии с той формой, в которой Максвелл первоначально записывал свои уравнения (за исключением того, что его плотность заряда в моих обозначениях выглядела бы как $c^{-2}\rho$). При другом выборе единиц множители, содержащие c, были бы распределены иначе.

12. Действительно, мы имеем *бесконечно* много x_i и p_i, но еще одно осложнение возникает в связи с тем, что мы не можем использовать непосредственно значения полей в этих координатах, поэтому для поля Максвелла нам необходимо ввести определенные «потенциалы», чтобы к нему можно было применить гамильтонову схему.

13. Т. е. не имеющие второй производной.

14. Уравнение Лоренца определяет *силу*, действующую на заряженную частицу со стороны электромагнитного поля, в котором та находится. Таким образом, если масса частицы известна, то второй закон Ньютона позволяет нам найти ускорение частицы. Но заряженные частицы часто движутся со скоростями, близкими к скорости света, так что начинают сказываться эффекты специальной теории относительности, для которых выбор массы частицы (см. следующий раздел) становится уже существенным. Именно по этой причине открытие правильного закона для силы, действующей на заряженную частицу, стало возможным только после появления на свет СТО.

15. Действительно, в некотором смысле, любая квантовомеханическая частица, встречающаяся в природе, сама по себе является часами. Как мы узнаем из главы 6, с любой квантовой частицей связано свое колебание, частота которого пропорциональна массе частицы (см. с. 211). Именно этот эффект позволил создать точнейшие современные (атомные и ядерные) часы.

16. «Излом» на мировой линии путешественника в точке B мог бы вызвать беспокойство у читателя: судя по картинке, путешественник в этой точке должен испытывать бесконечно большое ускорение. Но это несущественно. При конечном ускорении мировая линия путешественника будет иметь в точке B просто закругленный, или сглаженный изгиб, который очень слабо скажется на полном времени, которое путешественник проживает, и которое по-прежнему измеряется «длиной» (в смысле Минковского) всей его мировой линии.

17. С точки зрения наблюдателя **M**, эти пространства событий одновременны в смысле эйнштейновского *определения одновременности*, которое использует световые сигналы, посылаемые наблюдателем **M** и отражающиеся обратно к **M** из рассматриваемых точек пространства-времени. См., например, Риндлер [1982].

18. Это начальное значение *второй* производной по времени (или «ускорение») от формы. Быстрота изменения (или «скорость») формы первоначально считается равной нулю, так как сфера сначала находится в состоянии покоя.

19. Математическое описание этой переформулировки ньютоновской теории впервые было выполнено замечательным французским математиком Эли Картаном [1923], которое, разумеется, последовало после открытия общей теории относительности Эйнштейна.

20. Искривленные пространства — в том числе и многомерные — являющиеся в этом смысле локально евклидовыми, называются *римановыми многообразиями* в честь великого Бернгарда Римана (1826–1866), который первым исследовал такие пространства, опираясь в своих изысканиях на раннюю работу Гаусса, посвященную двумерному случаю. Здесь нам понадобится существенно модифицировать идеи Римана, вводя допущение о возможности замены локально евклидовой геометрии на геометрию *Минковского*. Такие пространства часто принято называть *лоренцевыми* многообразиями (принадлежащими к классу так называемых *псевдоримановых*, или, что менее логично, *полуримановых* многообразий).

21. Возможно, у читателя может возникнуть беспокойство по поводу того, каким образом это *нулевое* значение может быть *максимальным* значением «длины»! Но это именно так, хотя и в несколько бессодержательном смысле: геодезическая линия нулевой длины характеризуется тем, что не существует мировых линий *других* частиц, соединяющих (локально) любые две ее точки.

22. В действительности, это деление на эффекты деформации и изменения объема носит не настолько четкий характер, как я пытаюсь это изобразить. Тензор Риччи сам может дать определенный вклад в приливную деформацию. (Для световых лучей такое деление проводится однозначно; см. Пенроуз, Риндлер [1986], т. 2, глава 7.) Точное определение тензоров Вейля и Риччи см., например, в книге Пенроуза и Риндлера [1984], т. 1. (Герман Вейль (род. в Германии) был выдающимся математиком XX века, а Грегорио Риччи (род. в Италии) — весьма влиятельным геометром, создавшим также теорию тензоров.)

23. Правильная форма уравнений общей теории относительности была также найдена и Давидом Гильбертом (в ноябре 1915 года), однако все физические идеи, нашедшие отражение в этой теории, принадлежат исключительно Эйнштейну.

24. Для тех, кто разбирается в подобных вопросах, эти дифференциальные уравнения представляют собой полные *тождества Бьянки*, в которые подставлены уравнения Эйнштейна.

25. Существуют определенные (не очень убедительные) пути для обхода этого затруднения (см. Уилер, Фейнман [1945]).
26. Технически термин «гиперповерхность» более точен, чем «поверхность», так как объект не двумерен, а трехмерен.
27. Строгие *теоремы* на этот счет были бы очень полезны и интересны. Но пока их нет.
28. Невычислимый в рамках современной теории — которая дает (предварительно) достаточно бесполезный ответ: бесконечный!

Глава 6

Квантовая магия и квантовое таинство

Нужна ли философам квантовая теория?

Классическая физика — в полном согласии со здравым смыслом — рассматривает объективный мир, который существует «там, вовне». Этот мир эволюционирует ясным и детерминистским образом, управляемый точно сформулированными математическими уравнениями. Это также верно для теорий Максвелла и Эйнштейна, как и для исходной ньютоновской схемы. При этом считается, что физическая реальность существует независимо от нас самих, и как бы мы ни смотрели на классический мир — ничего в нем от этого не изменится. Кроме того, наше тело и наш головной мозг сами являются частью этого мира — а значит, эволюционируют в соответствии с теми же точными и детерминистскими классическими уравнениями. Все наши действия должны строго описываться этими уравнениями независимо от наших представлений о свободной сознательной воле, которой мы обладаем и которая может оказывать влияние на наше поведение.

Такая картина лежит, по-видимому, в основе самых серьезных[1] философских рассуждений по поводу природы реальности, нашего чувственного восприятия и нашей кажущейся свободы воли. Однако у некоторых возникает ощущение, что определенная роль должна быть отведена и квантовой теории — фундаментальной, но вызывающей смятение в умах картины мира, возникшей в первой четверти XX века, когда были обнаружены тончайшие расхождения между наблюдаемыми явлениями и их описаниями, которые предлагала классическая физика. У многих термин «квантовая теория» вызывает лишь смутные ассоциации с «принципом неопределенности», который говорит о невозможности точного описания системы на уровне частиц, атомов или молекул, позволяя использовать здесь лишь вероятностный подход. Как мы увидим в дальнейшем, квантовое описание является *точным*, хотя и радикально отличающимся от классического. Кроме того, мы обнаружим, что несмотря на общепринятое убеждение, вероятности *не возникают* на микроскопическом уровне (движение частиц, молекул и атомов происходит *детерминистично*), а появляются в результате некоторого загадочного крупномасштабного действия, ответственного за существование классического макромира, доступного нашим ощущениям. Мы должны попытаться понять это и выяснить, как квантовая теория изменяет наши взгляды на физическую реальность.

Можно было бы подумать, что квантовая теория вносит лишь незначительные поправки в описание физических явлений по сравнению с классической физикой. Но в действительности лишь благодаря этим поправкам могут существовать многие явления, происходящие в обычных масштабах. Само существование твердых тел, упругость и другие свойства материалов, химические свойства, цвет вещества, явления замерза-

ния и кипения, устойчивость наследственности — эти и многие другие знакомые нам явления невозможно объяснить без привлечения квантовой теории. Возможно, что и феномен сознания есть нечто, что нельзя объяснить, оставаясь в рамках классических представлений. Не исключено, что наш разум есть не просто элемент в игре так называемых «объектов» *классической* структуры, а скорее представляет собой качество, сущность которого коренится в необычных и удивительных особенностях физических законов, управляющих нашим миром. Пожалуй, что мы, как разумные существа, скорее должны были бы жить в квантовом, нежели в классическом мире, несмотря на все его богатство, разнообразие и удивительность. Возможно, что квантовый мир необходим, чтобы из обычного вещества можно было бы создать нас — чувствующих и мыслящих существ. Это — вопрос скорее к Богу, вознамерившемуся сотворить обитаемую вселенную, чем к нам! Но все это имеет непосредственное отношение и к нам. Если классический мир не есть нечто, частью чего могло бы быть наше сознание, то различия между классической и квантовой физикой должны каким-то образом влиять и на наш разум. К рассмотрению этой проблемы я еще вернусь позже.

Для того, чтобы основательно углубиться в философские вопросы и понять, *как ведет* себя наш мир и каково строение «разума», т. е. «нас самих», мы должны ближе познакомиться с квантовой теорией — самой точной и загадочной из физических теорий. Настанет время, когда наука достигнет *более* глубокого понимания природы, чем то, которое предлагает нам квантовая теория. Лично я склонен полагать, что квантовая механика есть лишь промежуточный и во многом еще неадекватный шаг на пути построения полной картины реального мира. Но это не освобождает нас от необходимости включения представлений квантовой теории в философскую картину реальности.

К сожалению, многие физики-теоретики придерживаются различных (но равноправных с точки зрения эксперимента) точек зрения относительно *актуальности* такой картины. Последователи Нильса Бора утверждают, что объективной картины реального мира *не существует*. С точки зрения квантовой теории «там, вовне» ничего *не существует*. Реальность же каким-то образом возникает только в связи с результатами «измерений». Согласно этой точке зрения квантовая теория представляет собой лишь вычислительную процедуру и не пытается описывать мир таким, каков он *есть* в действительности. Такое отношение к теории, на мой взгляд, является пораженческим, и я буду следовать позитивистскому способу рассмотрения, согласно которому *объективная физическая реальность* может быть описана квантовыми терминами: *квантовым состоянием*.

Существует точное уравнение — *уравнение Шредингера*, которое описывает полностью причинно обусловленную временную эволюцию этого состояния. Но взаимоотношение между изменяющимся во времени квантовым состоянием и наблюдаемым реальным миром происходит довольно странным образом. Время от времени — всякий раз, как только мы делаем заключение, что «измерение» уже произведено, мы вынуждены отказываться от того самого квантового состояния, за эволюцией которого мы наблюдали и использовать его только для вычисления вероятности, что оно скачком «перейдет» в одно из возможных *новых* состояний. В дополнение к странности этих «квантовых скачков» существует проблема того, какой должна быть физическая установка, позволяющая утверждать, что «измерение» действительно произведено. Измерительный прибор, в конечном счете, состоит из квантовых составляющих и поэтому должен эволюционировать в соответствии с уравнением Шредингера. Можно предположить, что и сами наблюдатели также построены из крохотных квантовых частиц. Является ли сознание необходимой составной частью процесса измерения? Думаю, что найдется немного физиков, готовых ответить положительно на этот вопрос.

Далее в этой главе мы рассмотрим некоторые необычные следствия этих «скачков» квантового состояния. Например, каким образом «измерение», производимое в одном месте, может вызвать «скачок» в другом удаленном месте! Но прежде нам необходимо познакомиться с еще одним необыч-

ным явлением. Допустим, что у объекта есть два различных, но совершенно равноправных маршрута движения. Если эти маршруты предоставлять ему по очереди, то он движется по ним одинаково хорошо. Но если открыть для него оба пути, то объект не может пройти *ни по одному* из них! Мы также рассмотрим более подробно, как реально описываются квантовые состояния, и увидим, насколько сильно это описание отличается от описания классических состояний. Например, частицы могут находиться сразу в двух местах! Мы обнаружим, насколько сложным становится квантовое описание системы многих частиц. Оказывается, что отдельная частица в такой системе не имеет определенного состояния. Только все вместе они обладают квантовым состоянием в виде сложной суперпозиции различных комбинаций друг друга. Мы увидим, как получается, что различные частицы одного типа не могут находиться в одинаковом квантовом состоянии. Мы подробно изучим необычное и сугубо квантовое свойство *спин*. Мы обсудим проблемы, возникающие в парадоксальном мысленном эксперименте с «кошкой Шредингера», и способы решения этой ключевой головоломки, предлагаемые разными теоретиками.

Возможно, что материал этой главы покажется читателю более сложным для восприятия и слишком специальным по сравнению с предыдущими и последующими главами. Но я не пытаюсь упростить изложение и вводить читателя в заблуждение, поэтому нам с вами предстоит поработать более серьезно, чем обычно. Это позволит нам приблизиться к подлинному пониманию квантового мира. В тех же случаях, когда изложение будет казаться вам непонятным, я советую проявить настойчивость и попытаться осознать картину в целом. Не следует отчаиваться, если вам так и не удастся достичь полного понимания, ибо трудности коренятся в самой природе излагаемого предмета!

Проблемы с классической теорией

Каким же образом выяснилось, что классическая физика не дает истинного описания нашего мира? Основной источник таких сведений — эксперимент. Квантовая теория не была всего лишь выдумкой теоретиков. Несмотря на огромное внутреннее сопротивление, они были вынуждены прийти к этому странному и во многом философски неудовлетворительному взгляду на окружающий нас мир. А произошло это потому, что классическая теория, несмотря на свое величие, столкнулась с серьезными трудностями. Главной из них было сосуществование физических объектов двух видов: *частиц*, описывающихся *конечным* числом параметров (шестью — тремя координатами и тремя компонентами импульсов), и *полей*, имеющих *бесконечно большое* число параметров. Такое деление в действительности оказывается физически непоследовательным. Для того, чтобы система частиц и полей пришла в состояние равновесия (или «полного покоя»), вся ее энергия должна перейти от частиц к полю. Это — проявление так называемого принципа «равномерного распределения энергии»: в равновесном состоянии вся энергия поровну распределяется между всеми степенями свободы системы. Так как поля обладают бесконечно большим числом степеней свободы, то на долю несчастных частиц вообще ничего не остается!

В этом случае классический атом был бы нестабилен, ибо движение его частиц полностью трансформировалось бы в волновые моды поля. Напомню, что в 1911 году британский физик-экспериментатор новозеландского происхождения Эрнест Резерфорд предложил модель атома, напоминающую солнечную систему. В центре такого атома подобно маленькому солнцу располагалось ядро, а вокруг него подобно планетам обращались электроны, удерживаемые на своих орбитах электромагнитными силами вместо гравитации. Фундаментальная и на первый взгляд неразрешимая проблема состояла в том, что в соответствии с уравнениями Максвелла электрон должен был за долю секунды упасть на ядро по спиральной траектории, непрерывно излучая при этом электромагнитные волны, интенсивность которых за такое малое время достигала бы бесконечной величины. Но ничего подобного не наблюдалось! То, что происходило *в действительности*, было

необъяснимо с точки зрения классической теории. Атомы могли излучать электромагнитные волны (свет) только определенного набора частот, в виде четких спектральных линий (рис. 6.1). Более того, эти частоты удовлетворяли «безумным» правилам [2], не имеющим под собой никакого основания в классической теории.

Одним из проявлений такой нестабильности системы полей и частиц стало явление, известное как «излучение абсолютно черного тела». Представьте себе объект, нагретый до определенной температуры, в котором электромагнитное излучение находится в тепловом равновесии с частицами. В 1900 году Рэлей и Джинс теоретически показали, что в этом случае вся энергия частиц должна быть без остатка «высосана» полем! Этот физически абсурдный результат получил название «ультрафиолетовой катастрофы», когда энергия безостановочно перетекает во все более и более высокочастотные колебания поля, в то время как в действительности природа никогда не ведет себя столь расточительно. Наблюдения показали, что энергия *низкочастотных* колебаний поля действительно соответствует предсказанию Рэлея и Джинса, но в *высокочастотной* части спектра (где ими была предсказана «ультрафиолетовая катастрофа») она *не возрастает* бесконечно, а спадает до нуля. Максимальное значение энергии при данной температуре приходится на определенную частоту или цвет (см. рис. 6.2). Хорошо знакомыми примерами этого могут служить красный цвет нагретой кочерги или желто-белый цвет раскаленного Солнца.

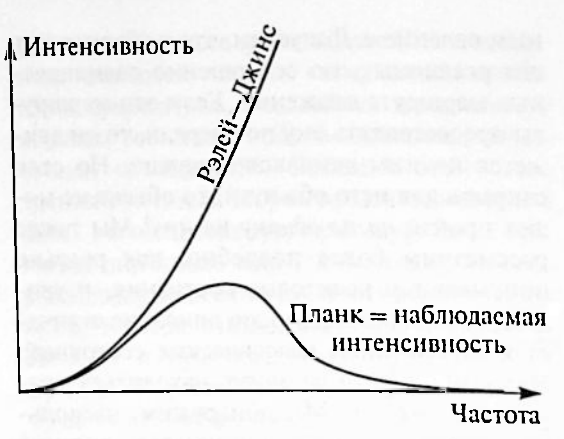


Рис. 6.2. Расхождение между интенсивностью излучения нагретого тела («абсолютно черного тела»), вычисленной в рамках классической теории, и наблюдаемой интенсивностью привели Планка к началам квантовой теории



Рис. 6.1. Атомы нагретого вещества испускают свет, который обычно имеет лишь очень определенные частоты. С помощью призмы различные частоты можно разделить и получить характерные для атомов спектральные линии

Начало квантовой теории

Как же разрешить все эти загадки? Очевидно, что исходную ньютоновскую схему частиц-корпускул *необходимо* дополнить максвелловским полем. Можно ли встать на противоположную точку зрения и предположить, что мир построен только из полей, а частицы представляют собой не что иное, как небольшие «сгустки» поля определенного вида? Этот подход имеет свои трудности, ибо такие частицы могли бы непрерывно изменять свою форму, извиваться и совершать колебания бесконечно большим числом способов. Но ничего подобного в действительности не наблюдается. В реальном мире все частицы одного вида, по-видимому, *идентичны*. Например, любые два электрона тождественны. Даже атомы и молекулы могут изменять свои конфигурации только дискретно [3]. Если частицы – это всего лишь поля, то необходимо ввести в теорию нечто новое, что заставило бы их иметь дискретные характеристики.

В 1990 году блестящий, но осторожный немецкий физик Макс Планк выдвинул революционную идею для подавления

высокочастотных мод излучения «абсолютно черного тела». Идея состояла в том, что излучение и поглощение электромагнитного поля может происходить только «квантами», энергия E которых связана с частотой ν следующим соотношением:

$$E = h\nu,$$

где h — новая фундаментальная постоянная природы, известная как *постоянная Планка*. Самое удивительное, что эта «бунтарская» идея позволила Планку достичь теоретического согласия с наблюдаемой зависимостью интенсивности излучения «абсолютно черного тела» от частоты (*закон излучения Планка*). (По современным данным постоянная Планка очень мала и составляет около $6{,}6 \times 10^{-34}$ Дж/с.) Смелая гипотеза Планка стала первым проблеском квантовой теории, но это событие не привлекло к себе внимания физиков до тех пор, пока Эйнштейн не выдвинул еще одну поразительную идею о том, что электромагнитное поле не только излучается, но и *существует* в виде таких дискретных порций. Таким образом, согласно Эйнштейну (и Ньютону, который высказывал аналогичное утверждение за два столетия раньше) свет представляет собой поток *частиц*! Вспомним, что в начале XIX века блестящий теоретик и экспериментатор Томас Юнг наглядно продемонстрировал волновую природу света, а Максвелл и Герц теоретически показали, что *свет* представляет собой колебания электромагнитного поля.

Каким образом свет может быть одновременно и частицами, и волнами? Ведь корпускулярная и волновая концепции представляются полностью противоположными. Тем не менее, одни экспериментальные факты явно указывают на то, что свет — это поток частиц, а другие на то, что свет — это волны. В 1923 году французский аристократ и проницательный физик маркиз Луи де Бройль продвинулся в этом вопросе еще дальше, высказав в своей докторской диссертации (которая снискала одобрение Эйнштейна!) идею о том, что частицы *материи* иногда ведут себя как волны! Частота ν волны де Бройля любой частицы с массой m также удовлетворяет соотношению Планка. Комбинируя это с формулой Эйнштейна $E = mc^2$, можно найти связь частоты ν с массой m:

$$h\nu = E = mc^2.$$

Таким образом, согласно идее де Бройля, раздельное существование частиц и полей, бывшее в почете у классической теории, *отвергается* природой! Действительно, все, что осциллирует с частотой ν, может существовать *только* в виде дискретных порций с массой $h\nu/c^2$. Природа каким-то образом «умудряется» построить непротиворечивый мир, в котором *частицы и осцилляции поля суть одно и то же*! Или, точнее, мир природы состоит из каких-то более тонких составляющих, а представления о «частице» и «волне» лишь частично отражают реальность.

Еще один яркий пример проявления соотношения Планка нашел в 1913 году Нильс Бор — датский физик и выдающийся мыслитель XX века. Правила Бора требовали, чтобы *угловой момент* (см. с. 254) электрона на ядерной орбите мог принимать только значения, кратные величине $h/2\pi$, для которой Дирак ввел более удобное обозначение $\hbar$:

$$\hbar = \frac{h}{2\pi}.$$

Таким образом, разрешены только следующие значения углового момента (относительно любой оси),

$$0,\ \hbar,\ 2\hbar,\ 3\hbar,\ 4\hbar,\ \ldots .$$

С учетом этого *нововведения* «планетарная» модель атома позволила с большой точностью вычислить частоты энергетических уровней и объяснить те «безумные» правила, которым *в действительности* следует природа.

Несмотря на поразительный успех, блестящая гипотеза Бора была только временной схемой, своего рода «новой заплатой на старые меха» и получила название «старой квантовой теории». Сегодняшняя квантовая физика произошла из двух независимых схем, предложенных позже немцем Вернером Гейзенбергом и австрийцем Эрвином Шредингером («матричной механики» в 1925 году и «волновой механики» в 1926 году, соответственно). Сначала две эти две схемы казались совершенно различными, но вскоре они были включены

в более общую теорию как ее эквивалентные представления. Это было сделано главным образом британским физиком-теоретиком Полем Адриеном Морисом Дираком. В последующих главах мы попытаемся окинуть беглым взглядом квантовую теорию и ее необычные следствия.

Эксперимент с двумя щелями

Рассмотрим «архетипичный» квантово-механический эксперимент, в котором пучок электронов, света или любых других «волн-частиц» направляется сквозь две узкие щели на расположенный позади них экран (рис. 6.3). Для большей конкретности выберем *свет* и условимся называть квант света «фотоном» согласно принятой терминологии. Наиболее очевидное проявление света как потока *частиц* (фотонов) наблюдается на экране. Свет достигает экрана в виде дискретных точечных порций энергии, которые всегда связаны с частотой света формулой Планка: $E = h\nu$. Энергия никогда не передается в виде «половинки» (или иной доли) фотона. Регистрация фотонов представляет собой явление типа «все или ничего». Всегда наблюдается только целое число фотонов.

Но при прохождении через две щели фотоны обнаруживают *волновое* поведение. Предположим, что сначала открыта только одна щель (а вторая — наглухо закрыта). Пройдя через эту щель, пучок света «рассеивается» (это явление называется *дифракцией* и является характерным для распространения волн). Пока еще можно при-

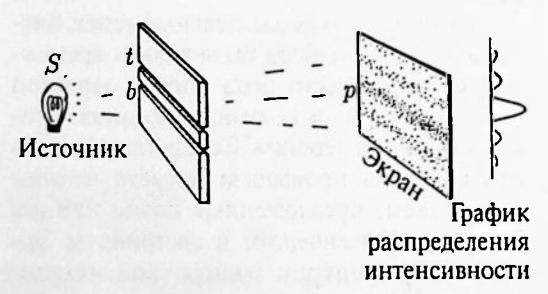


Рис. 6.3. Эксперимент с двумя щелями и монохроматическим светом (Обозначения на рисунке: *S* (англ. *sourse*) — источник, *t* (англ. *top*) — верхняя [щель], *b* (англ. *bottom*) — нижняя [щель]. — *Прим. ред.*)

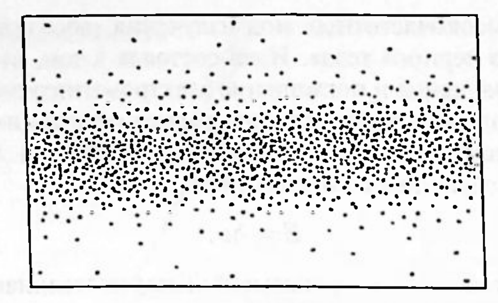


Рис. 6.4. Картина распределения интенсивности на экране, когда открыта только одна щель: наблюдается распределение дискретных крохотных пятнышек

держиваться корпускулярной точки зрения и считать, что расширение пучка обусловлено влиянием краев щели, заставляющем фотоны отклоняться на случайную величину в обе стороны. Когда свет, проходящий через щель, обладает достаточной интенсивностью (число фотонов велико), то освещенность экрана кажется равномерной. Но если интенсивность света уменьшить, то можно с уверенностью утверждать, что освещенность экрана распадется на отдельные пятна — в согласии с корпускулярной теорией. Яркие пятна располагаются там, где отдельные фотоны достигают экрана. Кажущееся равномерным распределение освещенности представляет собой статистический эффект, обусловленный очень большим числом участвующих в явлении фотонов (рис. 6.4). (Для сравнения, 60-ваттная электрическая лампа излучает около 100 000 000 000 000 000 000 фотонов в секунду!) При прохождении через щель фотоны действительно отклоняются случайным образом. Причем отклонения на различные углы имеют различные вероятности, что и порождает наблюдаемое распределение освещенности на экране.

Но главная трудность для корпускулярной картины возникает, когда мы открываем вторую щель! Предположим, что свет излучается желтой натриевой лампой, это значит, что он имеет чистый цвет без примеси, или, если воспользоваться физическим термином, свет *монохроматический*, т. е. имеет одну определенную частоту, или, на языке корпускулярной картины, все фотоны имеют одну и ту же энергию. Длина волны в данном случае составляет около 5×10^{-7} м. Предположим, что щели имеют в ширину

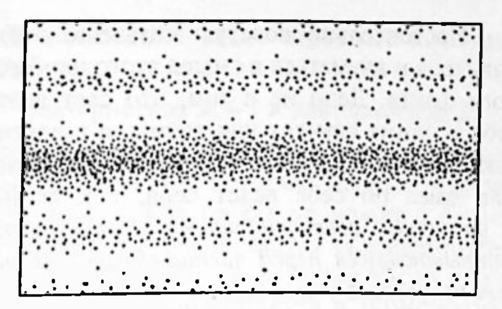


Рис. 6.5. Картина распределения интенсивности, когда открыты обе щели: наблюдается волнообразное распределение дискретных пятнышек

ну около 0,001 мм и отстоят друг от друга на расстояние около 0,15 мм, а экран находится от них на расстоянии около 1 м. При достаточно большой интенсивности света распределение освещенности все еще выглядит равномерным, но теперь в нем имеется некое подобие *волнообразности*, называемое *интерференционной картиной* — на экране примерно в 3 мм от центра наблюдаются полосы (рис. 6.5). Открывая вторую щель, мы надеялись увидеть вдвое бóльшую освещенность экрана (и это, действительно, было бы верно, если рассматривать *полную* освещенность экрана). Но оказалось, что теперь детальная *картина* освещенности полностью отлична от той, которая имела место при одной открытой щели. В тех точках экрана, где освещенность максимальна, его интенсивность оказывается не в *два*, а в *четыре* раза больше той, что была прежде. В других же точках, где освещенность минимальна, — интенсивность падает до нуля. Точки с нулевой интенсивностью, возможно, и представляют наибольшую загадку для корпускулярной точки зрения. Это те точки, которых фотон мог бы благополучно достичь, если бы открыта была только одна щель. Теперь же, когда мы открыли и вторую щель, неожиданно оказалось, что нечто *помешало* фотону попасть туда, куда он мог бы попасть прежде. Как могло случиться, что, предоставив фотону *альтернативный* маршрут, мы в действительности *воспрепятствовали* его прохождению по любому из маршрутов?

Если в качестве «размера» фотона принять длину его волны, то в масштабе фотона вторая щель находится от первой на расстоянии около 300 «размеров фотона» (а ширина каждой щели составляет около двух длин волн фотона) (рис. 6.6). Каким образом фотон, проходя через одну из щелей, «узнает» о том, открыта или закрыта другая щель? На самом деле, в принципе не существует предела для расстояния, на которое могут быть разнесены щели, для того, чтобы произошло явление «гашения или усиления».

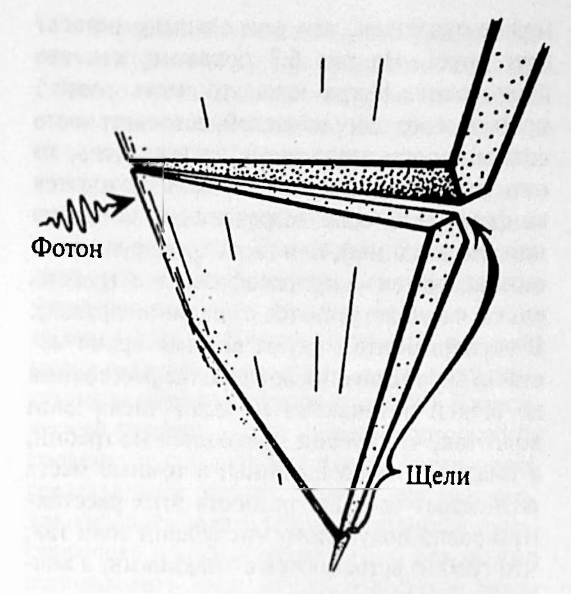


Рис. 6.6. Щели «с точки зрения» фотона! Разве может быть важно фотону, открыта или закрыта вторая щель, находящаяся на расстоянии около 300 «размеров фотона»?



Рис. 6.7. Чисто волновая картина позволяет нам осмыслить распределение светлых и темных полос на экране (но не дискретность) на языке интерференции волн

Создается впечатление, что когда свет проходит через одну или две щели, он ведет себя как *волна*, а не как корпускула (частица)! Такое гашение — *деструктивная интерференция* — хорошо известное свойство обычных волн. Если каждый из двух маршрутов порознь может быть пройден волной, то когда для нее открыты *оба* маршрута,

может оказаться, что они взаимно погасят друг друга. На рис. 6.7 показано, как это происходит. Когда какая-то часть волны, пройдя через одну из щелей, встречает часть волны, прошедшую через другую щель, то они усиливают друг друга, если находятся «в фазе» (т. е. если встречаются два гребня или две впадины), или гасят друг друга, если они находятся «в противофазе» (т. е. гребень одной части встречается с впадиной другой). В эксперименте с двумя щелями яркие места на экране возникают там, где расстояния до щелей отличаются на *целое* число длин волн так, что гребни приходятся на гребни, а впадины — на впадины, а темные места возникают там, где разность этих расстояний равна полуцелому числу длин волн так, что гребни встречаются с впадинами, а впадины — с гребнями.

Нет ничего загадочного в поведении обычной макроскопической классической волны, проходящей одновременно через две щели. Волна в конечном счете представляет собой всего лишь «возмущение» либо некоторой непрерывной среды (поля), либо некоторого вещества, состоящего из мириад крохотных точечных частиц. Возмущение может частично пройти через одну щель, частично через другую щель. Но в корпускулярной картине ситуация иная: каждый отдельный фотон сам по себе ведет себя, как волна! В некотором смысле каждая частица проходит *сразу через обе щели* и интерферирует *сама с собой*! Ибо, если значительно уменьшить полную интенсивность света, то можно гарантировать, что вблизи щелей будет находиться не более одного фотона одновременно. Явление деструктивной интерференции, когда два альтернативных маршрута каким-то образом «ухитряются» исключить друг друга из числа реализованных возможностей, есть нечто, применимое к *одному* фотону. Если для фотона открыт только один из двух маршрутов, то фотон может пройти по нему. Если открыт другой маршрут, то фотон может пройти второй вместо первого маршрута. Но если перед фотоном открыты *оба* маршрута, то эти две возможности чудесным образом исключают друг друга, и оказывается, что фотон не может пройти ни по одному из маршрутов!

Настоятельно советую читателю остановиться и вдуматься в смысл этого необычного факта. Дело не в том, что свет ведет себя в одних случаях как волны, а в других как частицы. *Каждая частица в отдельности* сама по себе ведет себя, как волна; и *различные альтернативные возможности, открывающиеся перед частицей, иногда могут полностью уничтожать друг друга*!

Действительно ли фотон расщепляется на два и частично проходит через одну щель, а частично — через другую? Большинство физиков будут возражать против такой постановки вопроса. По их мнению оба маршрута, открытых перед частицей, должны вносить вклад в конечный результат, они — всего лишь *дополнительные* способы движения, и не следует думать, будто частица должна расщепиться на две, чтобы пройти через щели. В подтверждение той точки зрения, что частица не проходит частично через одну щель и частично — через другую, можно рассмотреть видоизмененную ситуацию, в которой около одной из щелей помещен *детектор частиц*. В этом случае фотон (или любая другая частица) всегда появляется как единое целое, а не как некоторая доля целого: ведь наш детектор регистрирует либо целый фотон, либо полное отсутствие фотонов. Однако, если детектор расположен достаточно близко к одной из щелей, чтобы наблюдатель мог *различить*, через какую из них прошел фотон, то интерференционная картина на экране исчезает. Для того, чтобы имела место интерференция, по-видимому, необходимо «отсутствие знания» относительно того, через какую из щелей «действительно» прошла частица.

Чтобы получить интерференцию, *обе* альтернативы должны дать свой вклад, иногда «суммируясь», усиливая друг друга в два раза больше, чем можно было бы ожидать, а иногда «вычитаясь», чтобы загадочным образом *погасить* друг друга. Фактически же согласно правилам квантовой механики в действительности происходит нечто еще более загадочное! Конечно, альтернативы могут суммироваться (самые яркие точки на экране), альтернативы могут вычитаться (темные точки), но они также могут

образовывать и такие странные комбинации, как:

альтернатива $A + i \times$ альтернатива B,

где i — «квадратный корень из минус единицы» ($i = \sqrt{-1}$), с которым мы уже встречались в главе 3 (в точках на экране с промежуточной интенсивностью освещенности). В сущности *любое комплексное* число может играть роль коэффициента в «комбинации альтернатив»!

Возможно, читатель уже вспомнил высказанное мной в главе 3 предупреждение о том, что комплексные числа играют «абсолютно фундаментальную роль в структуре квантовой механики». Комплексные числа — не просто математические диковинки. Физиков вынудили обратить на них внимание убедительные и неожиданные экспериментальные факты. Чтобы понять квантовую механику, мы должны поближе познакомиться с языком комплекснозначных весовых коэффициентов. Давайте же рассмотрим, к каким это приводит последствиям.

Амплитуды вероятностей

Выбор фотона в приведенных выше рассуждениях не был продиктован ничем особенным. С тем же успехом для этого подошли бы электроны, любые другие частицы или даже целые атомы. Правила квантовой механики, насколько можно судить, утверждают, что и крикетные шары, и слоны должны вести себя описанным выше странным образом, где различные альтернативные возможности могут каким-то образом образовывать «суммы» состояний с комплексными весами! Однако нам никогда не приходилось реально *видеть* крикетные шары или слонов в виде столь странных «сумм». Почему? Это трудная и к тому же противоречивая тема, которую я не хотел бы сейчас затрагивать. А пока же мы просто допустим в качестве рабочего правила, что существуют два различных возможных уровня описания физической реальности, которые мы называем *квантовым уровнем* и *классическим уровнем*. Мы будем использовать эти странные комбинации состояний с комплекснозначными весами только на квантовом уровне. Крикетные же шары и слоны будут у нас объектами классического уровня.

Квантовый уровень — это уровень молекул, атомов и других субатомных частиц. Обычно считается, что это уровень явлений очень «малого масштаба», но эта «малость» не относится к физическим размерам. Мы увидим, что квантовые эффекты могут происходить на расстояниях многих метров или даже световых лет. Правильнее было бы считать, что нечто принадлежит «квантовому уровню», если это связано лишь с очень малыми изменениями энергии. (В дальнейшем я попытаюсь уточнить, о чем идет речь, главным образом в главе 8, с. 318.) Классический уровень — это «макроскопический» уровень, о котором мы имеем более непосредственные знания. Это — тот уровень, для которого верны наши обыденные представления о «происходящем», и где можно использовать наше обычное понятие вероятности. Мы увидим, что комплексные числа, которые нам приходится использовать на квантовом уровне, тесно связаны с классическими вероятностями. Но они не тождественны друг другу, и поэтому чтобы освоиться с этими комплексными числа, было бы очень полезно вспомнить для начала, как ведут себя классические вероятности.

Рассмотрим некую *неопределенную* классическую систему, то есть систему, о которой мы не знаем, в каком из двух альтернативных состояний A или B она находится. Такую систему можно было бы рассматривать как «взвешенную» комбинацию альтернатив A и B:

$p \times$ альтернатива $A + q \times$ альтернатива B,

где p — *вероятность* события A, а q — *вероятность* события B. (Напомним, что вероятность — действительное число, принимающее значение от 0 до 1. Вероятность 1 означает, что событие «заведомо произойдет», а вероятность 0 означает, что событие «заведомо не произойдет».) Если A и B — *единственно возможные* альтернативы, то сумма их вероятностей должна быть равна 1:

$$p + q = 1.$$

Если же существуют и другие возможности, то эта сумма должна быть меньше 1. В этом случае выражение $p : q$ дает *отношение* вероятности события A к вероятности события B. А сами вероятности событий A и B

(при условии, что имеются только эти две альтернативы) были бы равна, соответственно, $p/(p+q)$ и $q/(p+q)$. Мы может использовать такую интерпретацию и в том случае, когда сумма $p+q$ больше 1. (Такой способ вычисления вероятностей мог бы быть полезным, например, если бы мы многократно повторяли эксперимент, а p было бы количеством событий A, а q — количеством событий B). Мы будем говорить, что числа p и q *нормированы*, если $p+q=1$, в этом случае они дают сами вероятности, а не только отношения вероятностей.

Подобным образом мы поступаем и в квантовой физике, с тем лишь исключением, что в квантовой физике p и q — *комплексные* числа, в силу чего я предпочитаю их обозначить w и z, соответственно:

$$w \times \text{альтернатива } A + z \times \text{альтернатива } B.$$

Как же теперь нам истолковать w и z? Несомненно, что они не являются обычными вероятностями (или отношениями вероятностей), так как каждое из чисел w и z может по отдельности быть отрицательным или комплексным. Но во многих отношениях они ведут себя подобно вероятностям. Числа w и z (при соответствующей нормировке — см. далее) принято называть *амплитудами вероятности*, или просто *амплитудами*. Более того, часто используют терминологию, которая наводит на мысль о вероятностях, например: «Существует амплитуда w того, что произойдет событие A, и амплитуда z того, что произойдет событие B». Амплитуды еще не вероятности, но на миг попытаемся сделать вид, будто они являются вероятностями или, точнее, аналогами вероятностей на квантовом уровне.

Как проявляются *обычные* вероятности? Полезно представить себе какой-нибудь макроскопический объект, например, шарик, прошедший сквозь одну из двух щелей к стоящему позади экрану (как в описанном выше эксперименте с двумя щелями (см. рис. 6.3), но вместо прежнего фотона теперь фигурирует классический макроскопический шарик). Должна существовать некоторая вероятность $P(s,t)$ того, что отправившись из точки s шарик достигнет верхнего отверстия t, и некоторая вероятность $P(s,b)$ того, что шарик достигнет нижнего отверстия b. Кроме того, если мы выберем некоторую точку p на экране, то должна существовать некоторая вероятность $P(t,p)$ того, что шарик достигнет точки p на экране, пройдя через t, и некоторая вероятность $P(b,p)$ того, что он что шарик достигнет точки p, пройдя через b. Если открыто только отверстие t, то для того, чтобы найти вероятность того, что шарик действительно достигает точки p, пройдя через отверстие t, мы умножаем вероятность того, что он попадает из точки s в t, на вероятность того, что он попадает из t в точку p:

$$P(s,t) \times P(t,p).$$

Аналогично, если открыто только нижнее отверстие, то вероятность того, что шарик попадает из s в p, равна

$$P(s,b) \times P(b,p).$$

Если открыты *оба* отверстия, то вероятность того, что шарик попадает из s в точку p через t, по-прежнему равна первому произведению $P(s,t) \times P(t,p)$ (так, как если бы было открыто только отверстие t), и вероятность того, что шарик попадает из точки s в точку p через b, по-прежнему равна $P(s,b) \times P(b,p)$. Поэтому *полная* вероятность $P(s,p)$ того, что шарик, побывав в точке p, попадет в точку s, равна *сумме* двух приведенных выше вероятностей:

$$P(s,p) = P(s,t) \times P(t,p) + P(s,b) \times P(b,p).$$

На *квантовом* уровне эти правила остаются в точности такими же, с тем лишь исключением, что теперь роль вероятностей, с которыми мы имели дело в классическом случае, должны играть эти странные комплексные *амплитуды*. Например, в рассмотренном выше эксперименте с двумя щелями мы имеем амплитуду $A(s,t)$ того, что фотон достигнет верхней щели t из источника s, и амплитуду $A(t,p)$ того, что фотон достигнет точки p на экране из щели t, и, перемножив эти амплитуды, мы получим амплитуду

$$A(s,t) \times A(t,p)$$

того, что фотон достигнет точки p на экране через щель t. Как и в случае вероятностей, это — правильная амплитуда в предположении, что верхняя щель открыта независимо от того, открыта или не открыта нижняя

щель b. Аналогично, в предположении, что открыта нижняя щель b, мы получаем амплитуду

$$A(s,b) \times A(b,p)$$

того, что фотон достигнет точки p на экране через щель b (независимо от того, открыта или не открыта верхняя щель t). Если же открыты обе щели, то мы получаем полную амплитуду

$$A(s,p) = A(s,t) \times A(t,p) + A(s,b) \times A(b,p)$$

того, что фотон попадает в точку p из точки s.

Все это очень мило, но совершенно бесполезно, пока мы не знаем, как интерпретировать амплитуды, когда квантовый эффект увеличивается до классического уровня. Мы могли бы, например, поместить детектор фотонов, или *фотоячейку* в точке p, что дало бы нам способ увеличения события, происходящего на квантовом уровне, — прибытия фотона в точку p — до события, различимого на классическом уровне, скажем, громкого «щелчка». (С таким же успехом можно было бы взять в качестве экрана фотопластинку, на которой фотон оставляет видимое пятнышко, но для большей доходчивости мы все же воспользуемся фотоячейкой, издающей при срабатывании звуковой сигнал.) Должна существовать реальная *вероятность* того, что произойдет восприятие звукового «щелчка», а не одной из этих загадочных «амплитуд»! Как нам перейти от амплитуд к вероятностям, когда мы переходим с квантового уровня на классический? Оказывается, что для этого существует очень красивое, но удивительное правило.

Правило это состоит в том, что для получения классической вероятности, необходимо взять *квадрат модуля* квантовой комплексной амплитуды. Что такое «квадрат модуля»? Напомним как изображаются комплексные числа на плоскости Аргана (глава 3, с. 104). *Модуль* $|z|$ комплексного числа z есть просто расстояние от начала координат (т. е. от точки 0) до точки, изображающей число z. Квадрат модуля $|z|^2$ — просто квадрат этого числа. Таким образом, если

$$z = x + iy,$$

где x и y — действительные числа, то (по теореме Пифагора, так как отрезок прямой,

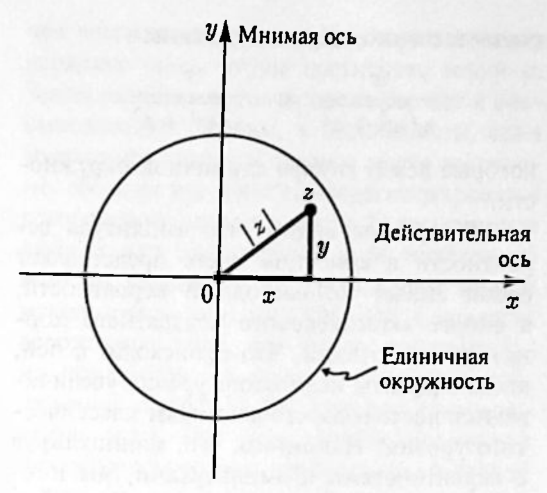


Рис. 6.8. Амплитуда вероятности представлена как точка z внутри единичной окружности на плоскости Аргана. Квадрат расстояния $|z|^2$ от центра может стать действительной вероятностью, если эффекты увеличены до классического уровня

соединяющий точки 0 и z, служит гипотенузой прямоугольного треугольника с катетами x и y) квадрат модуля равен

$$|z|^2 = x^2 + y^2.$$

Заметим, что для того, чтобы это выражение было настоящей «нормированной» вероятностью, значение $|z|^2$ должно быть заключено между 0 и 1. Это означает, что для того, чтобы быть надлежащим образом нормированной амплитудой, точка z на плоскости Аргана должна лежать где-то внутри *единичной окружности* (рис. 6.8). Однако иногда возникает необходимость рассматривать комбинации

$$w \times \text{альтернатива } A + z \times \text{альтернатива } B,$$

где w и z — всего лишь *пропорциональны* амплитудам вероятностей и поэтому не должны лежать внутри единичной окружности. Условие их *нормированности* (и, следовательно, того, что они дают настоящие амплитуды вероятностей) заключается в том, что сумма *квадратов* их *модулей* должна быть равна единице:

$$|w|^2 + |z|^2 = 1.$$

Если числа w и z не удовлетворяют этому условию нормировки, то настоящими амплитудами вероятностей альтернатив A и B,

соответственно, служат величины

$$\frac{w}{\sqrt{w^2+z^2}} \quad \text{и} \quad \frac{z}{\sqrt{w^2+z^2}},$$

которые лежат *внутри* единичной окружности.

Теперь мы видим, что амплитуда вероятности в конечном счете представляет собой аналог не настоящей вероятности, а скорее «комплексного квадратного корня» из вероятности. Что происходит с ней, когда эффекты квантового уровня увеличиваются настолько, что достигают классического уровня? Напомним, что, манипулируя с вероятностями и амплитудами, мы иногда сталкивались с необходимостью производить их умножение и сложение. Прежде всего заметим, что операция *умножения* не сопряжена с какими-либо проблемами при переходе от квантовых правил к классическим. Происходит это вследствие замечательного математического факта: квадрат модуля произведения двух комплексных чисел равен произведению квадратов модулей каждого из чисел:

$$|zw|^2 = |z|^2|w|^2.$$

(Это свойство непосредственно следует из геометрического смысла произведения двух комплексных чисел, приведенного в главе 3, но на языке действительной и мнимой частей $z = x+iy$, $w = u+iv$; это — прекрасное маленькое чудо. Проверьте сами!)

Из этого факта следует, что если в эксперименте с двумя щелями для частицы существует только один маршрут (открыта только одна щель, например t), то рассуждения можно строить «классически», и вероятности получатся одними и теми же, независимо от того, наблюдаем ли мы за прохождением частицы в промежуточных точках ее пути (в щели t)[1]. А квадраты модулей можно будет взять на любой стадии наших вычислений, например,

$$|A(s,t)|^2 \times |A(t,p)|^2 = |A(s,t) \times A(t,p)|^2.$$

Ответ — результирующая вероятность — получится одним и тем же.

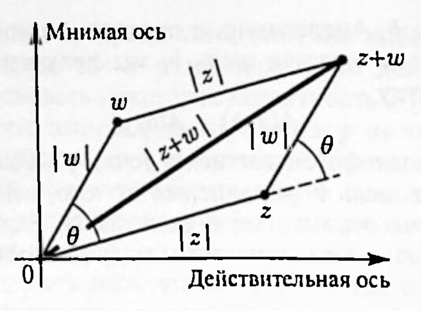


Рис. 6.9. Геометрический смысл добавочного члена $2|w||z|\cos\theta$ для квадрата модуля суммы двух амплитуд

Но если перед частицей открыт более чем один маршрут (например, если открыты обе щели), то необходимо образовывать *сумму*, и здесь-то и начинают обнаруживаться характерные особенности квантовой механики. Когда мы образуем квадрат модуля суммы $w+z$ двух комплексных чисел w и z, мы обычно *не получаем* только лишь сумму квадратов модулей этих чисел; существует дополнительный «поправочный член»:

$$|w+z|^2 = |w|^2 + |z|^2 + 2|w||z|\cos\theta,$$

где θ — угол, образуемый направлениями на точки z и w из начала координат на плоскости Аргана (рис. 6.9). (Напомним, что косинус угла есть отношение «прилежащий к углу катет/гипотенуза» для прямоугольного треугольника. Пытливый читатель, незнакомый с этой формулой, может попытаться самостоятельно вывести ее, используя геометрию, изложенную в главе 3. В сущности эта формула есть не что иное, как слегка «замаскированное» хорошо известное «правило косинуса»!) Именно поправочный член $2|w||z|\cos\theta$ описывает *квантовую интерференцию* между квантовомеханическими альтернативами. Значение $\cos\theta$ заключено между -1 и 1. При $\theta = 0°$ мы имеем $\cos\theta = 1$, и две альтернативы усиливают друг друга так, что полная вероятность оказывается больше суммы отдельных вероятностей. При $\theta = 180°$ мы имеем $\cos\theta = -1$, и две альтернативы стремятся погасить друг друга, в результате чего полная вероятность оказывается меньше

[1] Это наблюдение необходимо произвести так, чтобы не помешать прохождению частицы через щель t. Этого можно было бы достичь, разместив детекторы в *другом* месте — рядом с щелью s. Тогда можно будет делать *заключение* о прохождении частицы через щель t, когда эти детекторы *не срабатывают*!

суммы отдельных вероятностей (деструктивная интерференция). При $\theta = 90°$ мы имеем $\cos\theta = 0$, и получается ситуация, промежуточная между двумя упомянутыми выше: две вероятности просто суммируются. Для больших или сложных систем поправочные члены обычно «усредняются», так как «среднее» значение $\cos\theta$ равно нулю, и мы получаем обычные правила классической вероятности! Но на квантовом уровне эти члены описывают важные интерференционные эффекты.

Рассмотрим эксперимент с двумя щелями, когда обе щели открыты. Амплитуда того, что фотон достигает точки p, равна *сумме* $w + z$, где

$$w = A(s,t) \times A(t,p) \quad \text{и} \quad z = A(s,b) \times A(b,p).$$

В *самых ярких* точках экрана имеем: $w = z$ (так что $\cos\theta = 1$), откуда

$$|w + z|^2 = |2w|^2 = 4|w|^2,$$

что в 4 раза больше вероятности $|w|^2$, когда открыта только верхняя щель, и приводит к увеличению интенсивности потока большого числа фотонов в 4 раза, в полном согласии с экспериментом. В темных точках экрана имеем $w = -z$ (так что $\cos\theta = -1$), откуда

$$|w + z|^2 = |w - w|^2 = 0,$$

т. е. интенсивность равна *нулю* (деструктивная интерференция!) также в соответствии с наблюдением. Точно посередине между этими точками мы имеем: $w = iz$ или $w = -iz$ (так что $\cos\theta = 0$), откуда

$$|w + z|^2 = |w \pm iw|^2 = |w|^2 + |w|^2 = 2|w|^2,$$

что дает *вдвое бóльшую* интенсивность освещенности по сравнению с освещенностью только при одной щели (как в случае с классическими частицами). В конце следующего раздела мы узнаем, как рассчитывать, где именно расположены яркие, темные точки и точки с промежуточной интенсивностью освещенности.

И в заключение одно замечание. Когда открыты обе щели, амплитуда того, что частица достигнет точки p через щель t, в самом деле равна $w = A(s,t) \times A(t,p)$, но мы не можем интерпретировать квадрат ее модуля $|w|^2$ как вероятность того, что частица «действительно» прошла через верхнюю щель, чтобы достигнуть точки p. Такая интерпретация привела бы нас к бессмысленным ответам, в особенности, если точка p находится в темном месте на экране. Но если мы захотим «зарегистрировать» присутствие фотона в щели t, то усиливая эффект его присутствия (или отсутствия) *там* до классического уровня, мы *можем* использовать величину $|A(s,t)|^2$ в качестве вероятности того, что фотон действительно присутствует в щели t. Но такое наблюдение нарушило бы картину распределения волн. Для того, чтобы произошла интерференция, нам необходимо убедиться в том, что прохождение фотона через щели *остается на квантовом уровне*, так чтобы *оба* альтернативных маршрута давали свой вклад и иногда могли гасить друг друга. На квантовом уровне отдельные альтернативные маршруты обладают только амплитудами, но не вероятностями.

Квантовое состояние частицы

Как выглядит «физическая реальность» на квантовом уровне, где различные «альтернативные возможности», открытые перед системой, должны всегда обладать способностью сосуществовать, образуя суммы со странными комплекснозначными весами? Многие физики впадают в отчаяние при виде такой картины. Вместо этого они призывают рассматривать квантовую теорию только в качестве вычислительной процедуры для расчета вероятностей, а не объективной картины физического мира. Некоторые из них вполне серьезно заявляют, что квантовая теория проповедует невозможность получения объективной картины, по крайней мере той, которая согласуется с физическими фактами. Я же считаю такой пессимизм совершенно необоснованным. Во всяком случае было бы преждевременно на основании сказанного выше принять подобную точку зрения. Позднее мы рассмотрим некоторые из наиболее поразительных следствий квантовых эффектов, что возможно позволит нам понять причины такого отчаяния. Но пока давайте смотреть

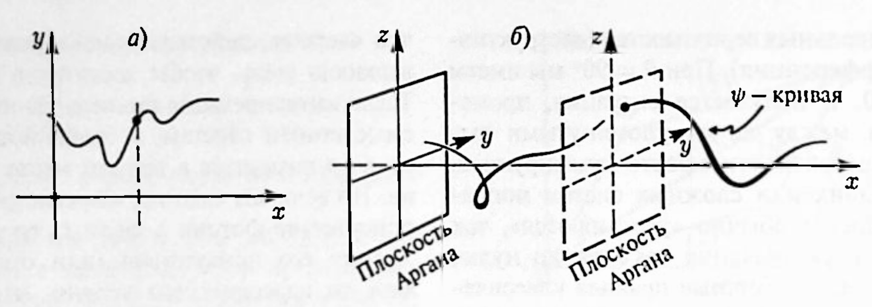


Рис. 6.10. *а*) График действительной функции действительной переменной x; *б*) график комплексной функции ψ действительной переменной x

на вещи более оптимистично и мужественно встретим все, что уготовила нам квантовая теория.

Первым предстанет перед нами *квантовое состояние*. Попытаемся мысленно представить себе одну-единственную квантовую частицу. Классически, частица определяется своим положением в пространстве, и для того, чтобы узнать, что произойдет с частицей дальше, нам также необходимо знать ее скорость (или, что эквивалентно, ее импульс). Квантовомеханически, *любое положение*, которое может занимать частица, является лишь одной их возможных «альтернатив» для частицы. Мы уже видели, что все альтернативы должны каким-то образом объединяться вместе с комплекснозначными весами. Набор этих комплекснозначных весов описывает квантовое состояние частицы. Обычно в квантовой теории принято использовать греческую букву ψ (произносится: «пси») для обозначения такого набора весов. Этот набор весов, рассматриваемый как комплекснозначная функция положения частицы, называется *волновой функцией* частицы. Для каждого положения x волновая функция принимает вполне определенное значение $\psi(x)$ — амплитуду вероятности того, что частица находится в положении x. Мы можем использовать одну букву ψ для обозначения квантового состояния как единого целого. Я разделяю ту точку зрения, что квантовое состояние ψ частицы — это и есть ее *физически реальное* положения в пространстве.

Каким же образом можно наглядно изобразить комплексную функцию ψ? Сделать это сразу для всего трехмерного пространства несколько затруднительно, поэтому мы немного упростим задачу и предположим, что наложенные связи позволяют частице двигаться только вдоль одномерной линии — например, оси x обычной (декартовой) системы координат. Если бы функция ψ была вещественной, то мы могли бы представить себе ось y, перпендикулярную оси x, и построить *график* функции ψ (рис. 6.10 *а*). Но в данном случае для изображения значения *комплексной* функции ψ нам требуется «комплексная ось y» — плоскость Аргана. Для этой цели вообразим, что мы можем использовать два других пространственных измерения: например, y-направление в качестве *действительной* оси плоскости Аргана, а z-направление — как *мнимую* ось. Для получения правильной картины волновой функции мы можем изобразить $\psi(x)$ (значение функции в точке x) точкой на этой плоскости Аргана (т. е. на плоскости yz, проходящей через каждую точку оси x). Когда положение точки x изменяется, то изменяется также и положение точки на плоскости Аргана. При этом точка описывает некоторую кривую в пространстве, извивающуюся вокруг оси x (рис. 6.10 *б*). Назовем эту кривую *ψ-кривой* рассматриваемой частицы. Если бы мы поместили в некоторой точке x детектор, то вероятность обнаружить частицу в данной точке можно найти, вычислив квадрат модуля амплитуды $\psi(x)$, т. е.

$$|\psi(x)|^2,$$

равный квадрату расстояния ψ-кривой от оси x [2)].

[2)] Здесь возникает техническая трудность, так как настоящая вероятность найти частицу *строго* в данной точке была бы равна нулю. Поэтому величину $|\psi(x)|^2$ мы предпочитаем называть *плотностью вероятности*.

Чтобы изобразить подобным образом волновую функцию, определенную на всем трехмерном физическом пространстве, понадобилось бы *пять* измерений: три — для физического пространства и два — для плоскости Аргана в каждой точке, в которой мы строим график функции $\psi(x)$. Однако наша упрощенная картина еще нам пригодится. Если мы захотим изучить поведение волновой функции вдоль произвольного направления в физическом пространстве, то для этого необходимо просто выбрать ось x вдоль этой линии, а два других пространственных измерения временно использовать в качестве действительной и мнимой осей на плоскости Аргана. Этот способ поможет нашему осмыслению эксперимента с двумя щелями.

Как я упоминал выше, в классической физике для того, чтобы определить, что будет происходить дальше, необходимо знать скорость (или импульс) частицы. В квантовой механике нам представляется значительная экономия. Волновая функция ψ *уже* содержит различные амплитуды для различных возможных импульсов! (Кое-кто из недовольных читателей может возразить, что «самое время» говорить об экономии, если принять во внимание, как сильно нам пришлось усложнить простую классическую картину точечной частицы. Хотя я во многом согласен с таким читателем, я все же советую не отвергать те лакомые кусочки, которые ему преподносят, ибо худшее еще впереди!) Каким образом амплитуды скоростей определяются волновой функцией ψ? На самом же деле лучше думать в терминах амплитуд импульсов. (Напомним, что импульс, или количество движения, равен скорости, умноженной на массу частицы, см. с. 254) Для этого следует применить к волновой функции ψ так называемый *гармонический анализ*. Подробно объяснять здесь, что это такое, было бы неуместно, скажу только, что он тесно связан с тем, что происходит с музыкальными звуками. Волну любой формы можно разложить в сумму различных «гармоник» (отсюда и термин «гармонический анализ»), которые представляют собой чистые тона различной высоты (т. е. с различными частотами). В случае волновой функции ψ «чистые тона» соответствуют различным возможным значениям импульса, которые может иметь частица, а величина вклада каждого «чистого тона» в ψ определяет амплитуду соответствующего значения импульса. Сами «чистые тона» называются *импульсными состояниями*.

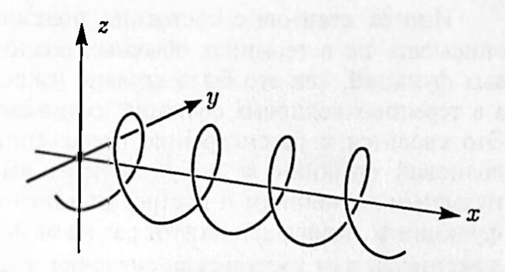


Рис. 6.11. Импульсное состояние имеет ψ-кривую в форме штопора

Как выглядит импульсное состояние, представленное ψ-функцией? Оно похоже на кривую, напоминающую по форме *штопор*, официальное математическое название которой — *винтовая линия* (рис. 6.11)[3]. Штопоры с частыми витками соответствуют большим импульсам, а штопоры, которые едва вращаются, — очень малым импульсам. Существует предельный случай, когда ψ-кривая вообще не делает витков и вырождается в прямую в случае нулевого импульса. В поведении винтовой линии неявно скрыто знаменитое *соотношение Планка*. Так как энергия E всегда пропорциональна частоте ν ($E = h\nu$), то частые витки означают короткую длину волны, большую *частоту* и, следовательно, большой импульс и высокую энергию, а редкие витки означают малую частоту и низкую энергию. Если плоскости Аргана ориентированы обычным способом (т. е. когда оси x, y, z образуют, как описано выше, правую тройку), то импульсы, направленные в положительном направлении оси x, соответствуют правым штопорам (которые обычно и используются).

Это означает, что на самом деле нам нужна вероятность найти частицу в некотором малом интервале фиксированных размеров. Таким образом, $\psi(x)$ определяет *плотность амплитуды*, а не просто амплитуду.

[3] На стандартном аналитическом языке любая из наших штопорообразных винтовых линий (т. е. любое импульсное состояние) задается формулой $\psi = e^{ipx/\hbar} = \cos(ipx/\hbar) + i\sin(ipx/\hbar)$ (см. главу 3, с. 116), где p — рассматриваемое значение импульса z.



Рис. 6.12. Дельта-функция в конфигурационном пространстве переходит в штопор в импульсном пространстве и наоборот

Иногда квантовые состояния полезно описывать не в терминах обычных волновых функций, как это было сделано выше, а в терминах волновых функций *импульсов*. Это сводится к рассмотрению разложения волновой функции ψ по различным импульсным состояниям и построению новой функции $\breve{\psi}$, зависящей на этот раз не от положения x, а от импульса p; значение $\breve{\psi}(p)$ при любом p задает величину вклада состояния с импульсом p в ψ-функцию. (Пространство величин p называется *импульсным пространством*.) Смысл $\breve{\psi}$ состоит в том, что при каждом конкретном выборе p комплексное число $\breve{\psi}(p)$ задает *амплитуду того, что частица имеет импульс p*.

Существует математическое название для соотношения между функциями ψ и $\breve{\psi}$. Каждая из этих функций называется *преобразованием Фурье* другой — в честь французского инженера и математика Жозефа Фурье (1768–1830). Я ограничусь здесь лишь несколькими замечаниями по поводу преобразования Фурье. Первое замечание: между ψ и $\breve{\psi}$ существует замечательная симметрия. Чтобы перейти от ψ назад к $\breve{\psi}$, мы по существу прибегаем к той же процедуре, которую использовали при переходе от ψ к $\breve{\psi}$. Теперь $\breve{\psi}$ становится объектом гармонического анализа. «Чистые тона» (т. е. штопоры в пространстве импульсов) на этот раз называются *конфигурационными состояниями*. Каждое положение x определяет такой «чистый тон» в пространстве импульсов, а величина такого вклада «чистого тона» в $\breve{\psi}$ дает значение $\psi(x)$.

Конфигурационное состояние соответствует (в терминах обычного пространства) некоторой функции ψ, имеющей острый пик в рассматриваемой точке x, а это значит, что все амплитуды равны нулю, за исключением амплитуды в данной точке. Такая функция называется *дельта-функцией* (Дирака), хотя, строго говоря, это — не совсем «функция» в обычном смысле, так как ее значение в точке x бесконечно велико. Аналогичным образом импульсные состояния (винтовые линии в конфигурационном пространстве) порождают дельта-функции в пространстве импульсов (рис. 6.12). Таким образом, оказывается, что преобразование Фурье винтовой линии есть дельта-функция и *наоборот*!

Описание в терминах конфигурационного пространства полезно всякий раз, когда требуется произвести измерение возможного положения частицы в пространстве, которое сводится к увеличению до классического уровня эффектов различных возможных положений частицы. (Грубо говоря, фотоэлементы и фотографические пластинки осуществляют измерение положения фотонов в пространстве.) Описание на языке импульсного пространства полезно, когда требуется измерить импульс частицы, т. е. увеличить до классического уровня эффекты различных возможных импульсов. (Эффекты отдачи или дифракции на кристаллах могут быть использованы для измерений импульса.) В каждом случае квадрат модуля соответствующей волновой функции (ψ или $\breve{\psi}$) дает искомую вероятность результата производимого измерения.

В заключение этого раздела обратимся еще раз к эксперименту с двумя щелями. Мы узнали, что согласно квантовой механике даже одна частица сама по себе должна обладать волновым поведением. Такая волна описывается волновой функцией ψ. Более всего похожи на волны волновые функции импульсных состояний. В эксперименте с двумя щелями мы рассматривали фотоны с определенной частотой; так что волновая функция фотона состояла из импульсных

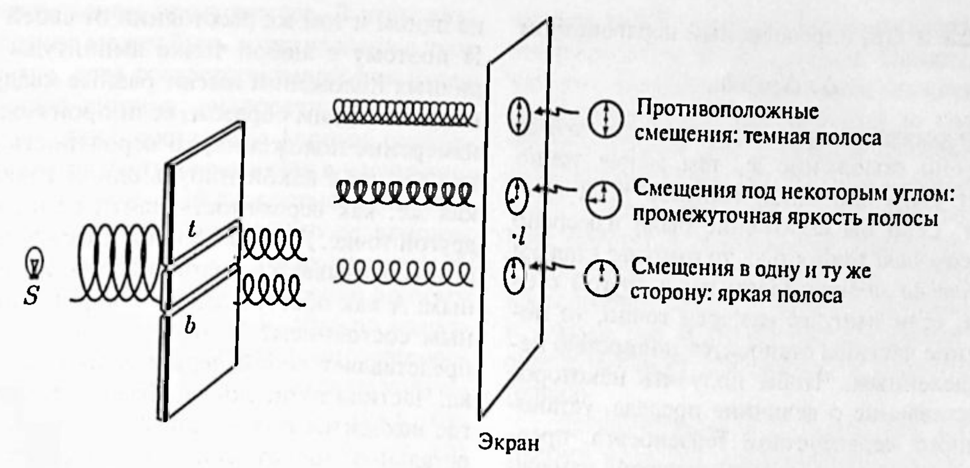


Рис. 6.13. Анализ эксперимента с двумя щелями в терминах штопорообразного представления импульсных состояний фотона

состояний различных направлений, в которых расстояние между соседними витками штопора — *длина волны* — было одно и то же на протяжении всей винтовой линии. (Длина волны определяется частотой.)

Волновая функция каждого фотона распространяется первоначально из источника в точке s и (если мы не следим за прохождением фотона через щели) проходит к экрану через обе щели. Однако только небольшая часть волновой функции проходит через щели, поэтому мы можем мысленно рассматривать щели как новые источники, каждый из которых по-отдельности испускает волновую функцию. Эти две части волновой функции интерферируют одна с другой так, что когда они доходят до экрана, в одних его точках они суммируются, а в других погашают друг друга. Чтобы выяснить, где волны суммируются и где гасят друг друга, выберем на экране некоторую точку l и рассмотрим прямые, проведенные к точке l от каждой из щелей t и b. Вдоль отрезка tl мы имеем одну винтовую линию, а вдоль отрезка bl — другую винтовую линию. (Мы также имеем винтовые линии вдоль линий st и sb, но если предположить, что источник находится на одном и том же расстоянии от обеих щелей, то на пути к щелям винтовые линии успеют совершить одинаковое число витков.) Число витков, которые винтовые линии совершат к тому моменту, когда они достигнут экран в точке l, зависит от длины отрезков tl и bl. Если эти длины отличаются на целое число длин волн, то в точке l винтовые линии окажутся совмещенными в *одном* направлении относительно своих осей (т. е. $\theta = 0°$, где θ определено в предыдущем разделе), так что соответствующие амплитуды сложатся и дадут *яркое* пятно. Если же эти линии отличаются по длине на целое число длин волн плюс половина длины волны, то в точке l винтовые линии окажутся совмещенными в *противоположных* направлениях относительно своих осей ($\theta = 180°$), поэтому соответствующие амплитуды погасят друг друга, и мы получим *темное* пятно. Во всех остальных случаях между смещениями винтовых линий в точке p образуется некоторый угол, поэтому соответствующие амплитуды будут суммироваться некоторым промежуточным образом, и мы получим пятно с промежуточной интенсивностью освещенности (рис. 6.13).

Принцип неопределенности

Большинству читателей приходилось слышать о *принципе неопределенности* Гейзенберга. Согласно этому принципу невозможно одновременно точно измерить (т. е. увеличить до классического уровня) положение и импульс частицы. Хуже того, существует *абсолютный предел* произведения погрешностей, с которыми могут быть измерены положение и импульс частицы, напри-

мер, Δx и Δp, определяемый неравенством

$$\Delta x \Delta p \geqslant \hbar.$$

Эта формула говорит нам, что чем точнее измерено положение x, тем менее точно может быть определен импульс p, и наоборот. Если бы положение было измерено с *бесконечной* точностью, то импульс стал бы *совершенно* неопределенным; с другой стороны, если импульс измерен точно, то положение частицы становится полностью неопределенным. Чтобы получить некоторое представление о величине предела, установленного неравенством Гейзенберга, предположим, что положение электрона измерено с погрешностью до нанометра (10^{-9} м), тогда его импульс стал бы настолько неопределенным, что уже через секунду после измерения бесполезно было бы искать электрон на расстоянии меньше 100 км от того места, где он находился в момент измерения!

Из описаний некоторых измерительных процессов создается впечатление, что это связано с некоторой неточностью, «встроенной» в сам процесс измерения. Согласно этой точке зрения, попытка локализовать электрон в вышерассмотренном эксперименте неизбежно сообщит ему случайный «толчок» такой интенсивности, что электрон, весьма возможно, улетит прочь с огромной скоростью, величина которой оговорена принципом неопределенности Гейзенберга. Из других же описаний мы узнаем, что неопределенность — свойство самой частицы, а ее движению присуща неизбежная случайность, которая означает, что поведение частицы непредсказуемо непосредственно на квантовом уровне. Есть и такие точки зрения, согласно которым квантовая частица есть нечто непостижимое, к чему неприменимы сами понятия классического положения и классического импульса. Ни один из этих подходов мне не нравится. Первый может ввести в заблуждение, второй заведомо неправилен, а третий излишне пессимистичен.

О чем в действительности говорит нам описание в терминах волновых функций? Прежде всего напомним наше определение импульсного состояния. Это тот случай, когда импульс известен точно. Кривая ψ имеет вид винтовой линии, всюду остающейся на одном и том же расстоянии от своей оси. И поэтому в любой точке амплитуды различных положений имеют равные квадраты модулей. Таким образом, если производится измерение положения, то вероятность найти частицу в какой-нибудь одной точке такая же, как вероятность найти ее в любой другой точке. Действительно, положение частицы оказывается полностью неопределенным! А как обстоит дело с конфигурационным состоянием? В этом случае ψ-кривая представляет собой дельта-функцию Дирака. Частица точно локализована в том месте, где находится пик дельта-функции, во всех остальных точках амплитуды равны нулю. Импульсные амплитуды лучше всего определять, перейдя в импульсное пространство. В этом случае их $\tilde{\psi}$-кривые имеют вид винтовых линий, так что амплитуды различных импульсов все имеют равные квадраты модулей. Результат измерения импульса частицы становится теперь совершенно неопределенным!

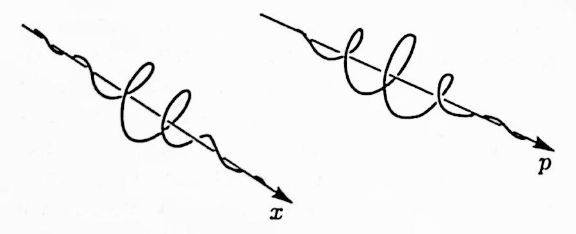


Рис. 6.14. Волновые пакеты, локализованные как в конфигурационном пространстве, так и в импульсном пространстве

Интересно рассмотреть промежуточный случай, когда координаты и импульсы отчасти ограничены, но только лишь в той степени, которая разрешена соотношением неопределенности Гейзенберга. Кривая ψ и соответствующая ей кривая $\tilde{\psi}$ (являющиеся Фурье-преобразованиями друг друга) для такого случая изображены на рис. 6.14. Обратите внимание на то, что расстояние от каждой из кривых до оси существенно отлично от нуля лишь в весьма малой области. Вдали от этой области кривые очень плотно прижимаются к оси. Это означает, что квадраты модуля заметно отличны от нуля только в очень ограниченной области как в конфигурационном пространстве, так

и в импульсном пространстве. В этом случае частица может быть локализована в пространстве, хотя соответствующий пик имеет некоторую ширину; аналогичным образом, импульс также достаточно хорошо определен, поэтому частица движется с достаточно хорошо определенной скоростью, а расплывание пика, характеризующего ее положение в пространстве, происходит не слишком быстро. Такое квантовое состояние принято называть *волновым пакетом*; обычно волновой пакет считается лучшим квантовотеоретическим приближением к классической частице. Однако из-за «размазанности» в значении импульса (т. е. скорости) следует, что волновой пакет со временем расплывается. И чем более он локализован в начальный момент времени в пространстве, тем быстрее он расплывается.

Эволюционные процедуры U и R

В приведенном выше описании временной эволюции волнового пакета неявно содержится *уравнение Шредингера*, которое говорит нам о том, как именно эволюционирует во времени волновой пакет. Действительно, уравнение Шредингера гласит, что каждая компонента разложения ψ по импульсным состояниям («чистым тонам») двигается со скоростью, равной величине c^2, деленной на скорость классической частицы, имеющей импульс данной компоненты. На самом деле, уравнение Шредингера математически сформулировано гораздо более лаконично. Мы обратимся к его точной записи несколько позднее. Оно по форме несколько напоминает уравнения Гамильтона или Максвелла (будучи тесно связано с обоими) и так же, как и эти уравнения, дает *полностью детерминистскую* эволюцию волновой функции, если волновая функция задана в какой-либо один момент времени (см. с. 254)!

Полагая, что ψ описывает мир в его «реальности», мы не обнаружим никакого индетерминизма, который, как предполагают некоторые, внутренне присущ квантовой теории, — не обнаружим, пока волновая функция ψ удовлетворяет детерминистской эволюции Шредингера. Будем называть это эволюционной U-процедурой. Однако всякий раз, когда мы «производим измерения», увеличивая квантовые эффекты до классического уровня, мы изменяем правила. Теперь вместо U мы используем совершенно другую процедуру, которую я обозначу R. Она состоит в образовании квадратов модулей квантовых амплитуд для получения классических вероятностей! [4] Именно эта и *только* эта R-процедура привносит неопределенности и вероятности в квантовую теорию.

Детерминистская U-процедура, по-видимому, является неотъемлемой частью той квантовой теории, на которой в основном сосредоточены помыслы активно работающих физиков; что же касается философов, то их больше интересует недетерминистская *редукция* R *вектора состояния* (или, как ее иногда называют более выразительно, *коллапс волновой функции*). Рассматриваем ли мы R просто как изменение «знания», которым мы располагаем о системе, или (как это делаю я) воспринимаем R как нечто «реальное», у нас имеется два совершенно *различных* математических подхода к описанию изменения во времени вектора состояния физической системы. В то время как U-процесс вполне детерминистский, R имеет вероятностный характер. U удовлетворяет комплексной квантовой суперпозиции состояний, а R грубо нарушает ее; U действует непрерывным образом, а R вопиющим образом разрывен. Исходя из стандартных процедур квантовой механики невозможно сделать заключение, что R-процесс может быть «выведен», как сложный случай U-процесса. R — это просто *другая*, отличная от U процедура, дающая вторую «половину» интерпретации квантового формализма. Весь индетерминизм квантовой теории происходит из R, а не из U. Но для изумительного согласия квантовой теории с наблюдательными фактами необходимы *оба* процесса: и U, и R.

Обратимся снова к волновой функции ψ. Предположим, что ψ описывает импульсное состояние. До тех пор, пока частица не взаимодействует с чем-нибудь, ψ благополучно остается импульсным состоянием до скончания времен. (Именно это

говорит нам уравнение Шредингера.) В любой момент времени, который мы выберем для «измерения импульса», мы получим один и тот же определенный ответ. Вероятностям здесь просто нет места. Предсказуемость остается здесь такой же четкой, как и в классической теории. Предположим, однако, что на некоторой стадии мы возьмемся измерить (т. е. увеличить до классического уровня) положение частицы. В этом случае мы получим целый массив амплитуд вероятности, модули которых нам предстоит возводить в квадрат. Имея такое изобилие вероятностей, мы столкнемся с полной неопределенностью в отношении того, каким будет результат измерения. Эта неопределенность согласуется с принципом неопределенности Гейзенберга.

С другой стороны, предположим, что мы начинаем с ψ, описывающей некоторое состояние частицы в конфигурационное пространстве. В этом случае согласно уравнению Шредингера ψ *не останется* в том же состоянии, а будет быстро расплываться. Тем не менее уравнение Шредингера полностью определяет, как происходит такое расплывание функции ψ. В ее поведении нет ничего недетерминистского или вероятностного. В принципе можно было бы предложить эксперименты, которые мы могли бы выполнить, чтобы проверить этот факт. (Подробнее об этом см. ниже.) Но если мы вдруг захотим измерить импульс, то получим амплитуды для всех различных возможных значений импульса, имеющие равные квадраты модулей, а результат эксперимента будет полностью неопределен — опять в полном соответствии с принципом неопределенности Гейзенберга.

Аналогичным образом, если исходить из ψ как волнового пакета, то его будущая эволюция полностью определяется уравнением Шредингера, и в принципе можно было бы предложить эксперименты, позволяющие проверить этот факт. Но как только мы вознамеримся произвести измерение над частицей каким-либо *другим* способом, например, измерить положение или импульс частицы, то мы сразу обнаружим, что неопределенности появляются снова (в соответствии с принципом неопределенности Гейзенберга) с вероятностями, задаваемыми квадратами модулей амплитуд.

Все это, несомненно, очень странно и таинственно. Но не означает, что мир непознаваем. В нарисованной мной картине мира многое подчиняется очень ясным и точным законам. Однако пока не существует ясного указания относительно того, когда следует прибегать к вероятностному правилу **R** вместо детерминистского правила **U**. Какой смысл следует вкладывать в выражение «выполнить измерение»? Почему (и когда) квадраты модулей амплитуд «становятся вероятностями»? Можно ли квантово-механически понять «классический уровень»? Это — глубокие и трудные вопросы, рассмотрением которых мы и займемся в следующей главе.

Одна частица — сразу в двух местах?

В приведенном выше описании я избрал гораздо более «реалистическую» точку зрения на волновую функцию, чем та, которая обычно принята среди квантовых физиков. Я придерживаюсь точки зрения, согласно которой «объективно реальное» состояние отдельной частицы действительно описывается ее волновой функцией ψ. Многие, видимо, находят такую позицию слишком трудной для того, чтобы ее можно было всерьез воспринимать. Одна из причин такого отношения, по-видимому, состоит в том, что эта позиция включает в себя представление об отдельных частицах как объектах, обладающих некоторой пространственной протяженностью, а не сосредоточенных в дискретных точках. Особую остроту эта ситуация приобретает для импульсного состояния, так как функция ψ распределена по всему пространству. Вместо того, чтобы представить себе частицу распределенной по всему пространству, люди предпочитают думать о ее положении как о «полностью неопределенном», так что все, что можно сказать о положении частицы, сводится к утверждению о том, что частица может находиться в каком-нибудь месте с такой же вероятностью, как и в любом другом. Однако мы видели, что волновая функция дает не только распределение вероятности различных

положений, но и распределение *амплитуд* для различных положений. Если мы знаем распределение амплитуд (т. е. функцию ψ), то (из уравнения Шредингера) мы также точно знаем, каким образом состояние частицы будет эволюционировать во времени. Представление о частице как об объекте, обладающем «пространственной протяженностью», необходимо нам для того, чтобы «движение» частицы (т. е. эволюция волновой функции ψ во времени) было определено таким образом. И если мы примем такое представление о частице, то движение ее станет точно определенным. «Вероятностная точка зрения» на ψ была бы уместной, если бы мы выполнили над частицей измерение ее положения, а $\psi(x)$ использовали бы далее только в форме квадрата ее модуля $|\psi(x)|^2$.

Похоже, что мы действительно должны согласиться с представлением о частице, как распределенной по обширным областям пространства и пребывающей в состоянии пространственной протяженности, пока не будет произведено следующее измерение ее положения. Даже будучи локализованной в конфигурационном пространстве, частица начинает в следующий момент времени обретать пространственную протяженность. Что касается импульсного состояния, то его, по-видимому, очень трудно принять в качестве «реальной» картины существования частицы, но еще труднее принять в качестве «реального» состояния с *двумя пиками*, которое имеет место, когда частица проходит через две щели (рис. 6.15). В вертикальном направлении форма волновой функции ψ имела бы два острых пика — по одному на каждой из щелей, являясь суммой[4)] волновой функции ψ_t, имеющей пик на верхней щели, и волновой функции ψ_b, имеющей пик на нижней щели:

$$\psi(x) = \psi_t(x) + \psi_b(x).$$

Если мы примем волновую функцию ψ как «реально» представляющую состояния частицы, то нам придется признать, что частица в самом деле *находится* в *двух* местах

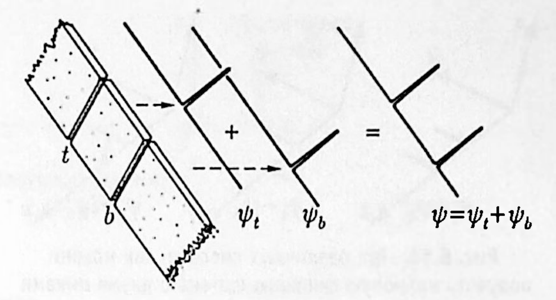


Рис. 6.15. Так как волновая функция фотона возникает от пары щелей, она имеет пики сразу в двух местах

одновременно! С этой точки зрения частица *реально прошла сразу через две щели.*

Стандартное возражение против утверждения о том, что частица реально «проходит сразу через две щели» сводится к следующему: если мы выполним измерение *на* щелях, чтобы определить, через какую из них прошла частица, то всегда обнаружим, что частица *целиком* проходит либо через одну, либо через другую щель. Но так происходит потому, что мы производим *измерение положения* частицы, поэтому ψ в этом случае дает только распределение вероятности $|\psi|^2$ для положения частицы — в соответствии с процедурой, основанной на вычислении квадрата модуля, и мы находим частицу либо в одном, либо в другом месте. Но существуют и другие типы измерений, которые *можно* производить на щелях, и эти измерения *отличны* от измерения положения. Для таких измерений нам необходимо было бы знать волновую функцию ψ с двумя пиками, а не только $|\psi|^2$, для различных положений x. При помощи таких измерений мы могли бы отличить состояние с двумя пиками

$$\psi = \psi_t + \psi_b,$$

приведенное выше, от других состояния с двумя пиками, таких, как

$$\psi_t - \psi_b$$

или

$$\psi_t + i\psi_b$$

(кривые ψ для каждого из этих различных случаев представлены на рис. 6.16). Так как измерения, различающие эти возможности, действительно существуют, они должны исчерпывать все *различные* возможные «реальные» способы существования фотона!

---

4) В более обычном квантовомеханическом описании эту сумму следовало бы разделить на нормирующий множитель, равный $\sqrt{2}$, т. е. взять сумму $(\psi_t + \psi_b)/\sqrt{2}$, но усложнять таким образом описание сейчас нет необходимости.

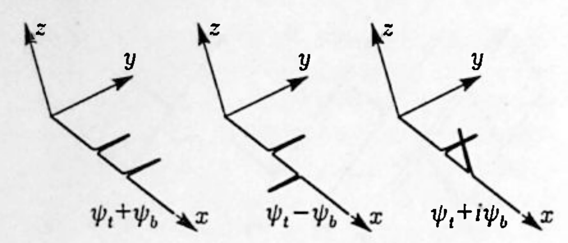


Рис. 6.16. Три различных способа, как можно получить волновую функцию фотона с двумя пиками

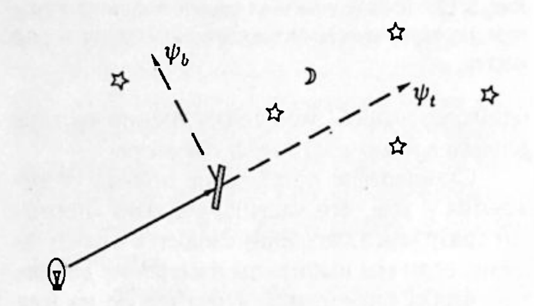


Рис. 6.17. Максимумы волновой функции с двумя пиками могут быть разнесены на расстояние в несколько световых лет. Этого можно достичь с помощью полупосеребренного зеркала

Щели не обязательно должны располагаться поблизости друг от друга для того, чтобы фотон мог пройти сквозь них одновременно. Чтобы понять, каким образом квантовая частица может находиться «в двух местах сразу» независимо от того, как далеко друг от друга расположены эти места, рассмотрим экспериментальную установку, немного отличающуюся от эксперимента с двумя щелями. Как и прежде, у нас имеется лампа, испускающая монохроматический свет, по одному фотону за раз; но вместо того, чтобы пропускать свет через две щели, отразим его от полупосеребренного зеркала, наклоненного к пучку под углом 45°. (Полупосеребренным называется зеркало, отражающее равно половину падающего на него света, тогда как вторая половина проходит прямо сквозь зеркало.) После встречи с зеркалом волновая функция фотона разделяется на две части, одна из которых отражается в сторону, а вторая продолжает распространяться в том же направлении, в котором первоначально двигался фотон. Как и в случае фотона, возникающего из двух щелей, волновая функция имеет два пика, но теперь эти пики разнесены на большее расстояние — один пик описывает отраженный фотон, другой — фотон, прошедший сквозь зеркало (рис. 6.17). Кроме того, со временем расстояние между пиками становится все больше и больше, увеличиваясь беспредельно. Представьте себе, что эти две части волновой функции уходят в пространство, и что мы ждем целый год. Тогда два пика волновой функции фотона окажутся на расстоянии светового года друг от друга. Каким-то образом фотон оказывается сразу в двух местах, разделенных расстоянием более чем в один световой год!

Есть ли какое-нибудь основание принимать такую картину всерьез? Разве мы не можем рассматривать фотон просто как некий объект, находящийся с вероятностью 50% в одном месте, и с вероятностью 50% — в другом! Нет, это невозможно! Независимо от того, как долго фотон находился в движении, всегда существует возможность того, что две части фотонного пучка могут быть отражены в обратном направлении и встретиться, в результате чего могут возникнуть интерференционные эффекты, которые не могли бы возникнуть из вероятностных весов двух альтернатив. Предположим, что каждая часть фотонного пучка встречает на своем пути полностью посеребренное зеркало, наклоненное под таким углом, чтобы свести обе части вместе, и что в точке встречи двух частей помещено еще одно полупосеребренное зеркало, наклоненное под таким же углом, как и первое зеркало. Пусть на прямых, вдоль которых распространяются части фотонного пучка, расположены два фотоэлемента (рис. 6.18). Что мы обнаружим? Если бы было справедливо, что фотон следует с вероятностью 50% по одному маршруту и с вероятностью 50% — по другому, то мы обнаружили бы, что оба детектора зафиксировали бы фотон каждый с вероятностью 50%. Однако в действительности происходит нечто иное. Если два альтернативных маршрута в точности равны по длине, то с вероятностью 100% фотон попадет в детектор A, расположенный на прямой, вдоль которой первоначально двигался фотон, и с вероятностью 0 — в любой другой детектор B. Иными словами фотон *с достоверностью* попадет в детектор A! (В этом можно убедиться, используя представление в форме винтовых

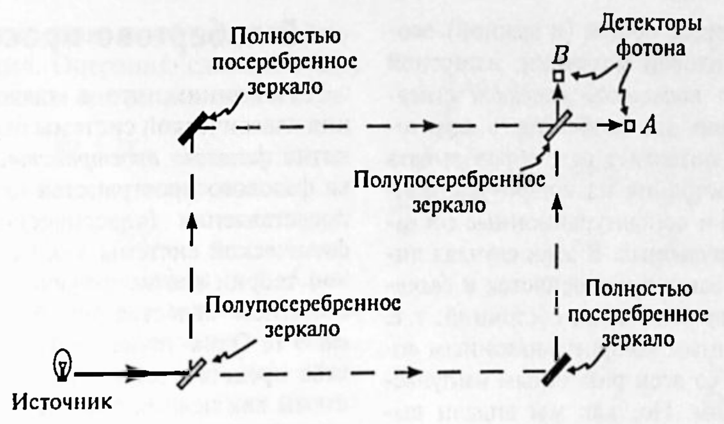


Рис. 6.18. Два пика волновой функции нельзя считать просто вероятностными весами локализации фотона в одном или другом месте. Два маршрута, избираемые фотоном, можно заставить интерферировать друг с другом

линий, приведенное выше для случая эксперимента с двумя щелями.)

Разумеется, такой эксперимент никогда не был поставлен для расстояний порядка светового года, но сформулированный выше результат не вызывает серьезных сомнений (у физиков, придерживающихся традиционной квантовой механики!) Эксперименты такого типа в действительности выполнялись для расстояний порядка многих метров или около того, и результаты оказывались в полном согласии с квантово-механическими предсказаниями (см. Уилер [1983]). Что же теперь можно сказать о *реальности* существования фотона между первой и последней встречей с полуотражающим зеркалом? Напрашивается неизбежным вывод, согласно которому фотон должен в некотором смысле *действительно* пройти оба маршрута сразу! Ибо если бы на пути любого из двух маршрута был помещен поглощающий экран, то вероятности попадания фотона в детектор A или B оказались бы одинаковыми! Но если открыты оба маршрута (оба одинаковой длины), то фотон может достичь только A. Блокировка одного из маршрутов *позволяет* фотону достичь детектора B! Если оба маршрута открыты, то фотон каким-то образом «знает», что попадание в детектор B *не разрешается*, и поэтому он вынужден следовать сразу по двум маршрутам.

Точка зрения Нильса Бора, согласно которой существованию фотона между моментами, когда производятся измерения, нельзя придать объективный «смысл», представляется мне слишком пессимистической относительно реальности состояния фотона. Квантовая механика дает нам *волновую функцию* для описания «реальности» положения фотона, и между полупосеребренными зеркалами волновая функция фотона как раз описывает состояние с двумя пиками, причем расстояние между пиками иногда бывает весьма значительным.

Заметим также, что утверждение «находится сразу в двух определенных местах» не полностью характеризует состояние фотона: нам необходимо отличать состояние $\psi_t + \psi_b$, например, от состояния $\psi_t - \psi_b$ (или, например, от состояния $\psi_t + i\psi_b$), где ψ_t и ψ_b теперь относятся к положениям фотона на каждом из двух маршрутов (соответственно «прошедшем» и «отраженном»!). Именно такого рода различие определяет, достигнет ли фотон с достоверностью детектора A, пройдя до второго полупосеребренного зеркала, либо он с достоверностью достигнет детектора B (или же он попадет в детекторы A и B с некоторой промежуточной вероятностью).

Эта загадочная особенность квантовой реальности, состоящая в том, что мы всерьез должны принимать во внимание, что частица может различными способами «находиться в двух местах сразу», проистекает из того, что нам приходится суммировать квантовые состояния, используя комплекснозначные веса для получения других квантовых состояний. Такого рода суперпозиция

состояний является общей (и важной) особенностью квантовой механики, известной под названием *квантовой линейной суперпозиции*. Именно эта особенность квантовой механики позволяет нам образовывать импульсные состояния из конфигурационных состояний и конфигурационные состояния — из импульсных. В этих случаях линейная суперпозиция применяется к *бесконечному* массиву различных состояний, т. е. ко всем различным конфигурационным состояниям или ко всем различным импульсным состояниям. Но, как мы видели выше, квантовая линейная суперпозиция весьма озадачивает, даже если мы применяем ее всего лишь к *двум* состояниям. По правилам квантовой механики *любые* два состояния, сколь бы сильно они ни отличались друг от друга, могут сосуществовать в любой комплексной линейной суперпозиции. Более того, любой объект, состоящий из отдельных частиц, должен обладать способностью существовать в такой суперпозиции пространственно далеко разнесенных состояний и тем самым «находиться в двух местах сразу»! В этом отношении формализм квантовой механики не проводит различия между отдельными частицами и сложными системами, состоящими из многих частиц. Почему же тогда мы не наблюдаем в повседневной жизни макроскопические тела, например, крикетные шары или даже людей, находящиеся в двух совершенно различных местах? Это — глубокий вопрос, и современная квантовая теория по сути дела не дает нам удовлетворительного ответа на него. В случае объекта, сравнимого с крикетным шаром, нам необходимо рассматривать систему на «классическом уровне». Или, как принято обычно говорить, производить «наблюдение» или «измерение» над крикетным шаром. Но в этом случае в качестве вероятностей, описывающих реальные альтернативы, необходимо рассматривать квадраты модулей комплекснозначных амплитуд вероятности, входящие в наши линейные суперпозиции в виде весов. Однако при этом сразу возникает сомнение в правомерности замены подобным способом квантовой **U**-процедуры на **R**-процедуру. К этому вопросу мы еще вернемся в дальнейшем.

Гильбертово пространство

Напомним, что в главе 5 для описания классической системы было введено понятие *фазового пространства*. Каждая точка фазового пространства используется для представления (классического) состояния физической системы как целого. В квантовой теории соответствующим аналогичным понятием является *гильбертово пространство*[5)]. Одна точка гильбертова пространства представляет *квантовое* состояние системы как целого. Нам необходимо бросить хотя бы беглый взгляд на математическую структуру гильбертова пространства. Надеюсь, что читателя не устрашит такая перспектива. В том, что я намереваюсь сказать, нет ничего математически очень сложного, хотя некоторые идеи могут показаться непривычными.

Наиболее фундаментальное свойство гильбертова пространства заключается в том, что оно представляет собой так называемое *векторное пространство*, а фактически комплексное векторное пространство. Это означает, что, *сложив* любые два элемента гильбертова пространства, мы получим элемент, также принадлежащий этому же пространству. Кроме того, когда мы производим сложение элементов гильбертова пространства, их разрешается умножать на комплекснозначные веса. Мы должны уметь делать такие операции, ибо они входят в состав только что рассмотренной *квантовой линейной суперпозиции*, а именно операции, ранее давшие нам фотонные состояния $\psi_t + \psi_b$, $\psi_t - \psi_b$, $\psi_t + i\psi_b$ и т. д. По существу, все что мы имеем в виду, используя термин «комплексное векторное пространство», сводится к разрешению образовывать взвешенные суммы указанного типа [5].

Удобно принять систему обозначений (предложенную главным образом Дираком), согласно которой элементы гильбертова пространства называются *векторами состояния* и обозначаются угловыми скобками $|\psi\rangle$, $|\chi|$, $|\varphi\rangle$, $|1\rangle$, $|2\rangle$, $|3\rangle$, $|n\rangle$, $|\uparrow\rangle$, $|\downarrow\rangle$, $|\rightarrow\rangle$, $|\nearrow\rangle$

5) Это важное понятие бесконечномерного пространства, с которым нам уже приходилось встречаться в предыдущих главах, ввел Давид Гильберт задолго до открытия квантовой механики и для совершенно других математических целей!

и т. д. Теперь эти символы обозначают квантовые состояния. Операцию сложения двух векторов состояния мы записываем в виде

$$|\psi\rangle + |\chi\rangle$$

или с комплексными весами w и z

$$w|\psi\rangle + z|\chi\rangle,$$

где $w|\psi\rangle$ означает $w \times |\psi\rangle$ и т. д. Соответствующим образом мы можем записать приведенные выше комбинации $\psi_t + \psi_b$, $\psi_t - \psi_b$, $\psi_t + i\psi_b$ в виде $|\psi_t\rangle + |\psi_b\rangle$, $|\psi_t\rangle - |\psi_b\rangle$, $|\psi_t\rangle + i|\psi_b\rangle$ и т. д. Мы можем также просто умножить *одно* состояние $|\psi\rangle$ на комплексное число w и получить

$$w|\psi\rangle$$

(в действительности это — частный случай приведенной выше комбинации состояний с комплексными весами при $z = 0$).

Напомним, что нам разрешается рассматривать комбинации с комплекснозначными весами w и z и в том случае, когда w и z — не являются амплитудами вероятности, а лишь им *пропорциональны*. Соответственно, мы принимаем правило, согласно которому весь вектор состояния можно умножить на отличное от нуля комплексное число, и физическое состояние от этого не изменится. (В результате такого умножения изменились бы значения весов w и z, но отношение $w : z$ осталось бы неизменным.) Каждый из векторов

$$|\psi\rangle,\ 2|\psi\rangle,\ -|\psi\rangle,\ i|\psi\rangle,\ \sqrt{2}|\psi\rangle,$$

$$\pi|\psi\rangle,\ (\lambda - 3i)|\psi\rangle \quad \text{и т. д.}$$

представляет *одно и то же* физическое состояние, как и любой вектор $z|\psi\rangle$, где $z \neq 0$. Единственный элемент гильбертова пространства, не допускающий интерпретацию как физическое состояние, есть нулевой вектор 0 (*начало координат* гильбертова пространства).

Чтобы получить некоторое геометрическое представление этой картины, рассмотрим сначала более привычное понятие «вещественного» вектора. Такой вектор принято изображать просто как *стрелку*, проведенную на плоскости или в трехмерном пространстве. Сложение двух таких векторов производится по правилу параллелограмма (рис. 6.19). Операция умножения

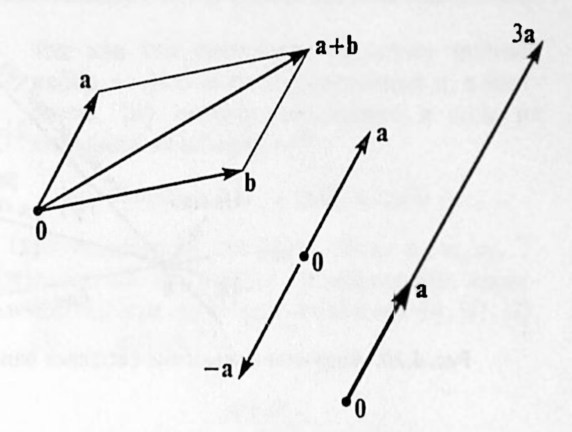


Рис. 6.19. Сложение и умножение на скаляры векторов в гильбертовом пространстве можно наглядно представить как соответствующие операции для векторов в обычном пространстве

вектора на положительное (вещественное) число сводится в таком представлении просто к умножению длины рассматриваемой стрелки на заданное число (направление стрелки при этом остается неизменным). Если же мы умножаем стрелку на отрицательное число, то направление стрелки изменяется на противоположное. Если число, на которое требуется умножить стрелку, равно 0, то мы получаем нулевой вектор **0**, который не имеет направления. (Вектор **0** представлен «нулевой стрелкой», имеющей нулевую длину.) Одним из примеров векторной величины может служить сила, действующая на частицу. Другими примерами могут служить классические скорости, ускорения и импульсы. Существуют также 4-векторы импульса, которые мы рассматривали в конце предыдущей главы. Это — векторы не в двумерном и не в трехмерном пространстве, а в четырехмерном. Но для гильбертова пространства нам понадобятся векторы с гораздо бо́льшим числом измерений (в действительности, часто даже бесконечномерные, но для нас это обстоятельство сейчас несущественно). Напомним, что мы всегда использовали стрелки, чтобы изобразить векторы в классическом фазовом пространстве, которое могло иметь очень высокую размерность. Говоря об «измерениях» фазового пространства, как и об «измерениях» гильбертова пространства, мы не имеем в виду обычные пространственные направления. Отнюдь! Каждое из-

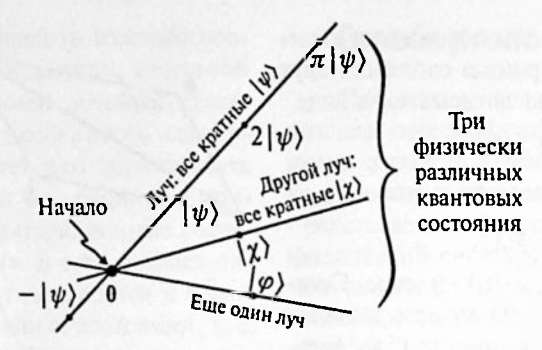


Рис. 6.20. Физические квантовые состояния описываются лучами в гильбертовом пространстве

мерение гильбертова пространства соответствует одному из различных независимых физических состояний квантовой системы.

Вследствие эквивалентности между $|\psi\rangle$ и $z|\psi\rangle$, физическое состояние в действительности соответствует целой *прямой*, *проходящей через начало координат* **0**, (или *лучу*) в гильбертовом пространстве (описываемом всеми кратными некоторого вектора), а не просто каким-то конкретным вектором, лежащим на этой прямой. Луч состоит из всех возможных кратных некоторого конкретного вектора состояния $|\psi\rangle$. (Следует иметь в виду, что речь идет о *комплексных* кратных, поэтому прямая в действительности представляет собой *комплексную* прямую, но об этом пока лучше не беспокоиться!) (См. рис. 6.20.) Скоро перед нами предстанет весьма изящная картина такого пространства лучей для случая *двумерного* гильбертова пространства. Другой предельный случай — бесконечномерное гильбертово пространство. Бесконечномерное гильбертово пространство возникает даже в простой ситуации локализации одной частицы. Тогда для каждого возможного положения, которое могла бы занимать частица, существует целое измерение! Каждое положение частицы определяет в гильбертовом пространстве целую «координатную ось», поэтому с учетом бесконечно многих различных положений частицы мы имеем бесконечно много различных независимых направлений (или «измерений») в гильбертовом пространстве. Импульсные состояния также могут быть представлены в *том же самом* гильбертовом пространстве. Поскольку импульсные состояния представимы в виде комбинаций конфигурационных состояний, то они соответствуют осям, идущим «по диагонали» — наклоненным относительно осей в конфигурационном пространстве. Совокупность всех импульсных состояний дает нам новую систему осей, и переход от осей конфигурационного пространства состояний к осям импульсного пространства состояний сводится к *повороту* в гильбертовом пространстве.

Не следует пытаться наглядно представить себе это сколько-нибудь точно. Такая попытка была бы неразумной! Однако некоторые идеи, почерпнутые из обычной евклидовой геометрии, могут оказаться очень полезными. В частности, рассматриваемые нами оси (*либо* все оси в конфигурационном пространстве состояний, *либо* все оси в импульсном пространстве состояний) следует считать взаимно *ортогональными*, т. е. расположенными под «прямыми» углами друг

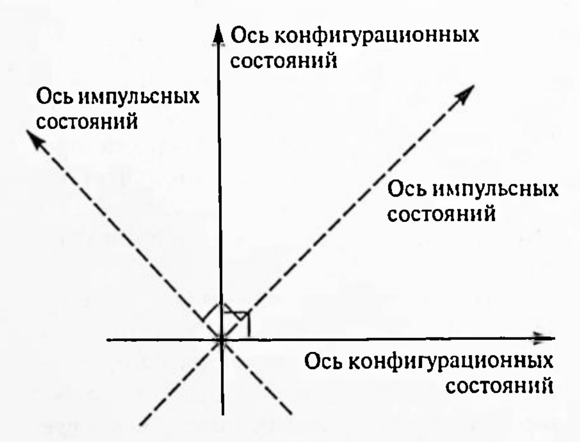


Рис. 6.21. Конфигурационные состояния и импульсные состояния приводят к различному выбору ортогональных осей в одном и том же гильбертовом пространстве

к другу. «Ортогональность» лучей — понятие, важное для квантовой механики. Ортогональные лучи соответствуют состояниям, которые *независимы* друг от друга. Различные возможные конфигурационные состояния частицы все взаимноортогональны, как и все различные возможные импульсные состояния. Но конфигурационные состояния не ортогональны импульсным состояниям. Весьма схематично эта ситуация представлена на рис. 6.21.

Измерения

Общее правило **R** для *измерения* (или наблюдения) требует, чтобы различные состояния квантовой системы, которые могут быть одновременно увеличены до классического уровня (на котором система должна выбрать одно из них), всегда должны быть взаимно *ортогональны*. Набор альтернатив, отобранный в результате *полного* измерения, образует систему ортогональных *базисных* векторов. Это означает, что каждый вектор в гильбертовом пространстве может быть (единственным образом) представлен в виде линейной комбинации этих векторов. Для измерения *положения*, произведенного над системой, состоящей из одной частицы, такие базисные векторы определяют те самые оси в конфигурационном пространстве состояний, о которых мы уже упоминали. Для измерения *импульса* это был бы другой набор, определяющий оси в импульсном пространстве состояний. Для полного измерения любого другого рода этот набор также был бы другим. После измерения состояние системы *скачком* переходит на одну из осей набора, соответствующего данному измерению, причем выбор оси происходит чисто случайным образом. Не существует динамического закона, который сказал бы нам, какая из осей будет выбрана природой. Ее выбор случаен, а значения вероятности определяются квадратами модулей амплитуд вероятности.

Предположим, что над системой, состояние которой $|\psi\rangle$, произведено некоторое полное измерение, причем базисом для выбранного измерения служит набор

$$|0\rangle, |1\rangle, |2\rangle, |3\rangle, \dots .$$

Так как эти состояния образуют полный набор, то любой вектор состояния и, в частности, $|\psi\rangle$ можно представить в виде их линейной комбинации[6)]

$$|\psi\rangle = z_0|0\rangle + z_1|1\rangle + z_2|2\rangle + z_3|3\rangle + \dots .$$

Геометрически коэффициенты $z_0, z_1, z_2, \dots$ являются *величинами ортогональных проекций* вектора $|\psi\rangle$ на различные оси $|0\rangle, |1\rangle, |2\rangle, |3\rangle, \dots$ (рис. 6.22).

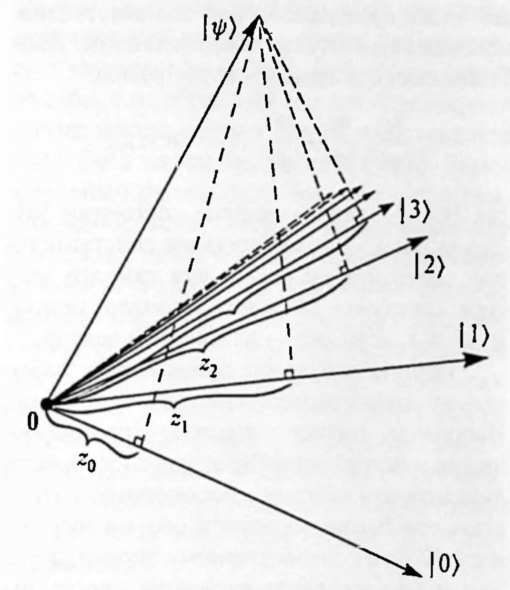


Рис. 6.22. Величины ортогональных проекций состояния $|\psi\rangle$ на оси $|0\rangle, |1\rangle, |2\rangle, \dots$ дают требуемые амплитуды $z_0, z_1, z_2, \dots$

Сразу возникает желание истолковать комплексные числа $z_0, z_1, z_2, \dots$ как искомые амплитуды вероятности, квадраты модулей которых давали бы различные вероятности того, что после измерения наша система будет находиться, соответственно, в состояниях $|0\rangle, |1\rangle, |2\rangle, |3\rangle, \dots$. Однако этого еще нельзя сделать, пока не определена «шкала» различных базисных векторов $|0\rangle, |1\rangle, |2\rangle, \dots$. Для этого мы должны оговорить, что в некотором смысле эти векторы

6) Не исключается также и случай, когда эта комбинация представляет собой *бесконечную* сумму векторов. *Полное* определение гильбертова пространства (которое, на мой взгляд, слишком формально для того, чтобы здесь вдаваться в его подробности) включает в себя правила, позволяющие оперировать с такими бесконечными суммами.

являются *единичными* (т. е. имеют единичную длину), и, таким образом, они образуют так называемый *ортонормированный* базис (элементы которого попарно *ортогональны* и *нормированы* на единицу) [6]. Если вектор $|\psi\rangle$ также нормирован на единицу, то искомые амплитуды действительно станут коэффициентами $z_0, z_1, z_2, \ldots$ вектора $|\psi\rangle$, а вероятности, которые требуется найти, будут равны $|z_0|^2, |z_1|^2, |z_2|^2, \ldots$. Если $|\psi\rangle$ — не единичный вектор, то приведенные выше числа *пропорциональны*, соответственно, искомым амплитудам и вероятностям. Действительные амплитуды будут равны

$$\frac{z_0}{|\psi|}, \quad \frac{z_1}{|\psi|}, \quad \frac{z_2}{|\psi|}, \quad \ldots \qquad \text{и т. д.,}$$

где $|\psi|$ — «длина» вектора состояния $|\psi\rangle$. Эта «длина» — положительное действительное число, определенное для каждого вектора состояния (0 имеет нулевую длину), и $|\psi| = 1$, если $|\psi\rangle$ — единичный вектор.

Полное измерение представляет собой весьма идеализированный тип измерения. Например, полное измерение положения частицы потребовало бы от нас способности локализовать частицу с бесконечной точностью, где бы во вселенной она ни находилась! К более элементарному типу измерения относится такое измерение, когда мы просто задаем вопрос типа «да или нет», например, такой: «Расположена ли частица справа (или слева) от некоторой прямой?» или «Лежит ли импульс частицы в некотором интервале?» и т. д. Измерения типа «да или нет» в действительности представляют собой наиболее фундаментальный тип измерения. (Например, используя только лишь измерения типа «да или нет», можно сколь угодно близко подойти к точному значению положения или импульса частицы.) Предположим, что результатом измерения типа «да или нет» оказывается **ДА**. Тогда вектор состояния должен находиться в области «ДА» гильбертова пространства, которую я обозначу **Y** [7]. С другой стороны, если результатом измерения типа «да или нет» оказывается **НЕТ**, то вектор состояния должен находиться в области «**НЕТ**» гильбертова пространства, которую я обозначу **N** [8]. Области **Y** и **N** полностью ортогональны друг другу в том смысле, что любой вектор состояния из области **Y** должен быть ортогонален любому вектору состояния из области **N** (и наоборот). Кроме того, любой вектор состояния $|\psi\rangle$ может быть (единственным образом) представлен в виде суммы векторов, принадлежащих каждой из областей **Y** и **N**. Если воспользоваться математической терминологией, то можно сказать, что области **Y** и **N** являются *ортогональными дополнениями* друг друга. Таким образом, $|\psi\rangle$ однозначно представим в виде

$$|\psi\rangle = |\psi_Y\rangle + |\psi_N\rangle,$$

где $|\psi_Y\rangle$ принадлежит **Y**, а $|\psi_N\rangle$ принадлежит **N**. Здесь $|\psi_Y\rangle$ означает *ортогональную проекцию* состояния $|\psi\rangle$ на **Y**, а $|\psi_N\rangle$ — ортогональную проекцию состояния $|\psi\rangle$ на **N** (рис. 6.23).

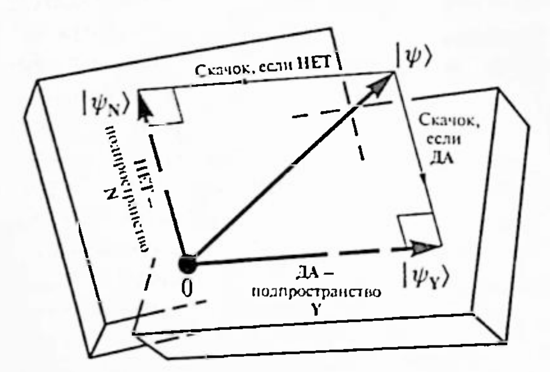


Рис. 6.23. Редукция вектора-состояния. Измерение может быть описано в терминах пары подпространств **Y** и **N**, каждое из которых является ортогональным дополнением другого. После измерения состояние $|\psi\rangle$ скачком переходит в свою проекцию на одно из этих подпространств с вероятностью, задаваемой множителем, показывающим, во сколько раз квадрат длины вектора состояния уменьшается при переходе к проекции

Если результат измерения есть **ДА**, то $|\psi\rangle$ скачком переходит в $|\psi_Y\rangle$, а если результат есть **НЕТ**, то в $|\psi_N\rangle$. Если вектор состояния $|\psi\rangle$ нормирован, то соответствующие вероятности того и другого исхода равны *квадратам длин*

$$|\psi_Y|^2 \quad \text{и} \quad |\psi_N|^2$$

[7] От английского *yes* — «да». — *Прим. ред.*

[8] От английского *no* — «нет». — *Прим. ред.*

состояний-проекций. Если же вектор $|\psi\rangle$ не нормирован, то каждый из этих квадратов необходимо разделить на $|\psi|^2$. (По «теореме Пифагора» $|\psi|^2 = |\psi_Y|^2 + |\psi_N|^2$, т. е. сумма вероятностей, как и должно быть, равна единице!) Заметим, что вероятность скачкообразного перехода состояния $|\psi\rangle$ в состояние $|\psi_Y\rangle$ определяется отношением, показывающим, во сколько раз квадрат длины вектора $|\psi\rangle$ уменьшается при таком проецировании.

В заключение необходимо сделать одно замечание относительно таких «актов измерения», которые можно производить над квантовой системой. Из самих основ квантовой теории следует, что для *любого* состояния, скажем, для $|\chi\rangle$, существует измерение типа «да или нет» [7], результатом которого будет **ДА**, если измеряемое состояние пропорционально $|\chi\rangle$, и **НЕТ**, если оно ортогонально $|\chi\rangle$. Таким образом, введенная выше область **Y** могла бы состоять из всех состояний, кратных любому выбранному состоянию $|\chi\rangle$. Из этого утверждения, по-видимому, следует весьма сильное заключение о том, что векторы состояния должны быть *объективно реальными*. Каким бы ни было состояние физической системы (давайте назовем его $|\chi\rangle$), существует в принципе выполнимое измерение, для которого $|\chi\rangle$ — *единственное* (с точностью до пропорциональности) состояние, *с достоверностью* приводящее к результату ДА. Может оказаться, что для некоторых состояний $|\chi\rangle$ выполнить такое измерение будет чрезвычайно трудно, а порою практически «невозможно». Но тот факт, что согласно теории существует *принципиальная* возможность такого измерения, приведет позднее в этой главе к некоторым поразительным следствиям.

Спин и сфера Римана состояний

Величину, которую в квантовой механике принято называть *«спином»*, иногда считают самой «квантовомеханической» из всех физических величин, поэтому мы поступим разумно, уделив ей некоторое внимание. Что такое спин? По существу, спин — это мера, характеризующая вращение частицы. Термин «спин» [9] действительно наводит на мысль о чем-то, напоминающем вращение крикетного шара или бейсбольного мяча. Вспомним понятие *углового момента*, который, подобно энергии и импульсу, является *сохраняющейся* величиной (см. главу 5, с. 160, а также с. 211). Угловой момент тела остается постоянным во времени до тех пор, пока движение тела не возмущает трение или какие-нибудь другие силы. Он и есть то, чем на самом деле является квантовомеханический спин, но сейчас нас интересует «вращение» *отдельной* частицы самой по себе, а не обращение по орбитам мириад частиц вокруг общего центра масс (как это было бы в случае крикетного шара). Замечательный физический факт состоит в том, что большинство частиц, обнаруживаемых в природе, действительно совершают «вращение» в только что указанном смысле, причем каждая частица обладает спином, величина которого специфична только для нее [8]. Но, как мы увидим дальше, спин отдельной квантовомеханической частицы обладает некоторыми весьма экстравагантными свойствами, — совсем не теми, которые мы могли бы ожидать, исходя из своего опыта обращения с закрученным крикетными шарами.

Прежде всего, для частиц определенного типа *величина* спина всегда *одна и та же*. Изменяться (причем очень странным образом, о чем мы вскоре узнаем) может только направление спина. Это резко контрастирует с крикетным шаром, который может быть закручен всеми возможными способами как угодно сильно или слабо в зависимости от того, как он был запущен! Для электрона, протона или нейтрона величина спина всегда равна $\hbar/2$, т. е. ровно *половине* наименьшего положительного значения, которое по Бору было изначально допустимым для квантованной величины углового момента атомов. (Напомним, что допустимыми значениями были $0, \hbar, 2\hbar, 3\hbar, \ldots$.) Здесь же нам требуется половина фундаментальной единицы $\hbar$, и, в некотором смысле, $\hbar/2$ сама по себе есть даже более фундаментальная единица. Такая величина углового

[9] От английского *spin* — «вращение». — *Прим. ред.*

момента не была бы допустима для объекта, состоящего только из орбитальных частиц, не вращающихся самих по себе. Такая величина может возникнуть только потому, что спин — это *внутренне присущее* свойство самой частицы (т. е. он не является результатом орбитального движения ее «частей» вокруг некоторого центра).

Частица со спином, равным *нечетному* кратному $\hbar/2$ (т. е. $\hbar/2$, $3\hbar/2$ или $5\hbar/2$ и т. д.) называется *фермионом* и обладает любопытной квантовомеханической особенностью: полный поворот на 360° переводит ее вектор состояния не в себя, а в себя со знаком *минус*! Многие частицы, встречающиеся в природе, относятся к числу фермионов, и мы еще узнаем позднее о них и их необычных свойствах, столь жизненно важных для нашего существования. Остальные частицы со спином, равным *четному* кратному $\hbar/2$, т. е. целому кратному $\hbar$ (а именно $0, \hbar, 2\hbar, 3\hbar, \ldots$), называются *бозонами*. При повороте на 360° вектор состояния бозона переходит точно *в себя*.

Рассмотрим частицу *с половинным спином*, т. е. со значением спина $\hbar/2$. Для определенности я буду называть такую частицу *электроном*, но ею с таким же успехом мог бы быть протон или нейтрон, а также атом подходящего вида. («Частица» может состоять из отдельных частей, если ее можно рассматривать квантовомеханически как единое целое с вполне определенным полным угловым моментом.) Предположим, что наш электрон покоится, и рассмотрим только его спиновое состояние. Пространство квантовомеханических состояний (гильбертово пространство) оказывается в этом случае *двумерным*, поэтому мы можем выбрать базис, состоящий всего лишь из *двух* состояний. Я обозначу их $|\uparrow\rangle$ и $|\downarrow\rangle$, чтобы указать, что в состоянии $|\uparrow\rangle$ спин вращается слева направо относительно вертикального направления *снизу вверх*, в то время как в состоянии $|\downarrow\rangle$ спин вращается слева направо относительно вертикального направления *сверху вниз* (рис. 6.24). Состояния $|\uparrow\rangle$ и $|\downarrow\rangle$ взаимно ортогональны, и мы считаем их нормализованными ($|\uparrow|^2 + |\downarrow|^2 = 1$). Любое возможное состояние спина электрона представимо в виде линейной суперпозиции, например, $w|\uparrow\rangle + z|\downarrow\rangle$, именно этих *двух* ортонормированных состояний $|\uparrow\rangle$ и $|\downarrow\rangle$, т. е. состояний *спин вверх* и *спин вниз*.

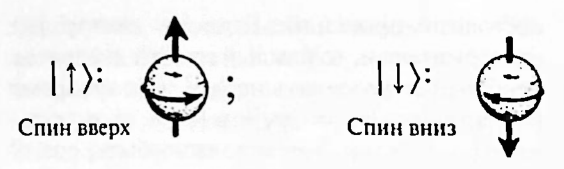


Рис. 6.24. Базис спиновых состояний электрона состоит всего лишь из двух состояний. В качестве них принято выбирать состояния *спин вверх* и *спин вниз*

Нужно сказать, что в состояниях *спин вверх* и *спин вниз* нет ничего особенного. С тем же успехом мы могли бы описывать спин, вращающийся слева направо вокруг любого другого направления, например, *слева-направо* $|\rightarrow\rangle$ и противоположного ему *справа-налево* $|\leftarrow\rangle$. Тогда (при подходящем выборе комплексных весов) мы получили бы для $|\uparrow\rangle$ и $|\downarrow\rangle$ [10]:

$$|\rightarrow\rangle = |\uparrow\rangle + |\downarrow\rangle \quad \text{и} \quad |\leftarrow\rangle = |\uparrow\rangle - |\downarrow\rangle.$$

Это позволяет нам по-новому взглянуть на ситуацию. Любое спиновое состояние электрона есть линейная суперпозиция двух ортогональных состояний $|\rightarrow\rangle$ и $|\leftarrow\rangle$, т. е. спинов *направо* и *налево*. Можно выбрать какое-нибудь совершенно произвольное направление, например, вектор состояния $|\nearrow\rangle$. Он также является линейной комбинацией спинов $|\uparrow\rangle$ и $|\downarrow\rangle$ с некоторыми комплексными коэффициентами, скажем,

$$|\nearrow\rangle = w|\uparrow\rangle + z|\downarrow\rangle,$$

а любое спиновое состояние было бы представимо в виде линейной комбинации этого состояния $|\nearrow\rangle$ и ортогонального ему [9] состояния $|\swarrow\rangle$. (Заметим, что понятие «ортогональный» в гильбертовом пространстве не обязательно означает «образующий прямой угол с...» в обычном пространстве. Ортогональные вектора состояния в гильбертовом пространстве в данном случае соответствуют диаметрально противоположным направлениям, а не образующим друг с другом прямой угол.)

Каково геометрическое соотношение между направлением в пространстве, опре-

---

[10] Здесь и выше я предпочел не загромождать формулы множителями типа $1/\sqrt{2}$, которые нужны, если мы требуем, чтобы векторы $|\rightarrow\rangle$ и $|\leftarrow\rangle$ были нормированными.

деляемым спином $|\nearrow\rangle$, и двумя комплексными числами w и z? Так как физическое состояние, задаваемое спином $|\nearrow\rangle$, останется неизменным, если мы умножим $|\nearrow\rangle$ на любое ненулевое комплексное число, то значение имеет только *отношение* числа z к числу w. Обозначим это отношение через

$$q = \frac{z}{w}.$$

Тогда q будет обычным комплексным числом за исключением того, что теперь ему разрешено принимать значение $q = \infty$, чтобы не упускать из рассмотрения ситуацию с $w = 0$, т. е. когда спин направлен вертикально вниз. Если $q \neq \infty$, то мы можем представить q как точку на плоскости Аргана, как мы делали это в главе 3. Представим себе, что эта плоскость Аргана расположена горизонтально в пространстве, причем действительная ось направлена вправо в вышеуказанном смысле (т. е. в направлении спинового состояния $|\rightarrow\rangle$). Представим теперь сферу единичного радиуса, центр которой совпадает с началом координат плоскости Аргана, а точки $1, i, -1, -i$ лежат на экваторе этой сферы. Рассмотрим точку, совпадающую с южным полюсом этой сферы, который мы обозначим ∞. Осуществляя проекцию из южного полюса, мы отобразим всю плоскость Аргана на нашу единичную

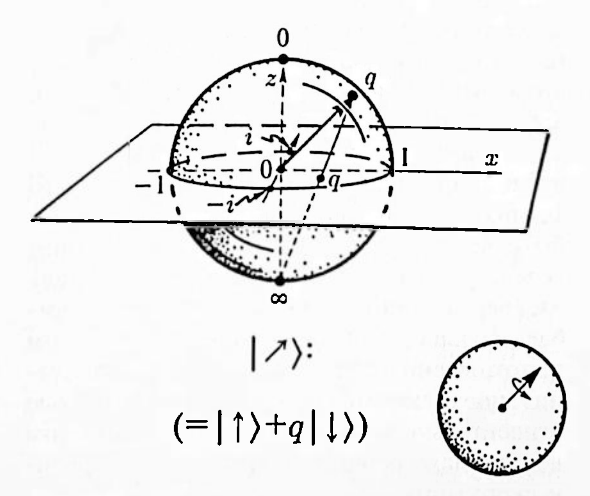


Рис. 6.25. Сфера Римана, представленная как пространство физически различных спиновых состояний частицы со спином 1/2. Сфера Римана стереографически спроецирована из ее южного полюса (∞) на плоскость Аргана, проходящую через экватор сферы

сферу. В результате любая точка q на плоскости Аргана окажется поставленной в соответствие единственной точке q на этой сфере, лежащей на прямой, соединяющей эти две точки с южным полюсом (рис. 6.25). Такое соответствие называется *стереографической проекцией* и обладает многими красивыми геометрическими свойствами (например, сохраняет углы и отображает окружности в окружности). Такая проекция позволяет нам параметризовать точки сферы комплексными числами вместе с ∞, т. е. множеством возможных комплексных отношений q. Сфера, параметризованная таким образом, называется *сферой Римана.* Геометрический смысл сферы Римана для спиновых состояний электрона состоит в том, что направление спина, задаваемое соотношением $|\nearrow\rangle = w|\uparrow\rangle + z|\downarrow\rangle$, определяется реальным направлением из центра в точку $q = z/w$, как показано на изображении сферы Римана. Заметим, что северный полюс соответствует состоянию $|\uparrow\rangle$, задаваемому соотношением $z = 0$, т. е. $q = 0$, а южный полюс — состоянию $|\downarrow\rangle$, задаваемому соотношением $w = 0$, т. е. $q = \infty$. Самая правая точка сферы Римана помечена значением $q = 1$, что соответствует состоянию $|\rightarrow\rangle = |\uparrow\rangle + |\downarrow\rangle$, а самая левая точка сферы Римана соответствует $q = -1$, что дает спиновое состояние $|\leftarrow\rangle = |\uparrow\rangle - |\downarrow\rangle$. Самая дальняя задняя точка сферы Римана помечена значением $q = i$, соответствующим состоянию $|\uparrow\rangle + i|\downarrow\rangle$, в котором спин направлен прямо от нас, а самая близкая точка сферы Римана помечена значением $q = -i$, соответствующим состоянию $|\uparrow\rangle - i|\downarrow\rangle$, в котором спин направлен прямо к нам. Произвольная точка, помеченная q, соответствует состоянию $|\uparrow\rangle + q|\downarrow\rangle$.

Как все это связано с измерением, которое можно было бы произвести над спином электрона? [10] Выберем некоторое направление в пространстве и обозначим его α. Если мы измеряем спин электрона в этом направлении, то ответ ДА означает, что электрон (теперь) действительно вращается слева направо вокруг направления α, в то время как ответ **НЕТ** означает, что электрон вращается слева направо вокруг направления, противоположного α.

Предположим, что мы получили ответ ДА, и обозначим результирующее состояние $|\alpha\rangle$. Если мы просто повторим измерение, используя в точности такое же направление α, как прежде, то с вероятностью 100% обнаружим, что ответ будет ДА. Но если при втором измерении мы изменим направление и выберем новое направление β, то обнаружим, что вероятность ответа ДА (состояние перепрыгивает в $|\beta\rangle$) будет несколько меньшей, и существует некоторая возможность появления во втором измерении ответа НЕТ (состояние перепрыгивает в направление, противоположное β). Как нам вычислить эту вероятность? Ответ на этот вопрос содержится в предписаниях, приведенных в конце предыдущего раздела. Вероятность ответа ДА для второго измерения оказывается равной

$$\frac{1}{2}\left(1+\cos\vartheta\right),$$

где ϑ — угол между направлениями[11] α и β. Соответственно, вероятность ответа НЕТ для второго измерения равна

$$\frac{1}{2}\left(1-\cos\vartheta\right).$$

Отсюда видно, что если второе измерение производится под прямым углом к первому, то вероятность составляет 50% в обоих случаях ($\cos 90^\circ = 0$); результат второго измерения полностью случаен! Если угол между двумя измерениями острый, то ответ ДА более вероятен, чем ответ НЕТ. Если этот угол — тупой, то ответ НЕТ более вероятен, чем ДА. В предельном случае, когда направление β противоположно направлению α, вероятность равна 0 для ответа ДА и 100% для ответа НЕТ, т. е. результат второго измерения заведомо обратен результату первого измерения. (См. Фейнман и др. [1965] для дальнейшего знакомства со спином.)

Сфера Римана действительно играет фундаментальную (но не всегда признанную) роль в *любой* квантовой системе с двумя состояниями, описывая (с точностью до коэффициента пропорциональности) набор возможных квантовых состояний. Для частицы с полуцелым спином ее геометрическая роль особенно очевидна, так как точки сферы соответствуют возможным пространственным направлениям спиновых осей. Увидеть роль сферы Римана во многих других ситуациях труднее. Рассмотрим фотон, только что прошедший через две щели или отразившийся от полупосеребренного зеркала. Состояние фотона есть некоторая линейная комбинация типа $|\psi_t\rangle + |\psi_b\rangle$, $|\psi_t\rangle - |\psi_b\rangle$ или $|\psi_t\rangle + i|\psi_b\rangle$ двух состояний $|\psi_t\rangle$ и $|\psi_b\rangle$, описывающих две совершенно различные локализации. Сфера Римана по-прежнему описывает набор физически различных возможностей, но теперь лишь *абстрактно*. Состояние $|\psi_t\rangle$ представлено северным полюсом («верхушкой») сферы, а состояние $|\psi_b\rangle$ — южным полюсом («дном») сферы. Соответственно, состояния $|\psi_t\rangle + |\psi_b\rangle$, $|\psi_t\rangle - |\psi_b\rangle$ и $|\psi_t\rangle + i|\psi_b\rangle$ представлены различными точками на экваторе, и в общем случае состояние $w|\psi_t\rangle + z|\psi_b\rangle$ представлено точкой, задаваемой отношением $q = z/w$. Во многих случаях (как и в рассматриваемом примере) возможности «богатства сферы Римана» довольно глубоко упрятаны, не имея прямого отношения к геометрии пространства!

Объективность и измеримость квантовых состояний

Несмотря на то, что мы обычно располагаем только вероятностями для результата некоторого эксперимента, нам кажется, что в квантовомеханическом состоянии есть все же нечто *объективное*. Часто высказывают утверждение, что векторы состояния — всего лишь удобное представление «нашего знания» о физической системе — или, может быть, вектор состояния описывает на самом деле не одну-единственную систему, а лишь дает вероятностную информацию об «ансамбле» большого числа одинаковым образом приготовленных систем. Такие высказывания поражают меня неразумной робостью относительно того, что квантовая механика должна нам сообщить о «реальности» физического мира.

Некоторая осторожность и сомнение относительно «физической реальности» векторов состояния, по-видимому, проистекает из того, что согласно теории набор измеримых величин строго ограничен. Рассмотрим

спиновое состояние электрона, как было описано выше. Предположим, что спиновым состоянием оказывается $|\alpha\rangle$, но мы этого не знаем, т. е. нам неизвестно «направление» α, вокруг которого как вокруг оси вращается электрон. Можем ли мы определить это направление с помощью эксперимента? Нет, не можем. Лучшее, что мы можем сделать, это извлечь «один бит» информации, т. е. получить ответ на один вопрос типа «да или нет». Мы можем выбрать в пространстве некоторое направление β и измерить спин электрона в этом направлении. В результате измерения мы получим ответ либо **ДА**, либо **НЕТ**, но после этого информация о первоначальном направлении спина будет утрачена. Получив ответ **ДА**, мы будем знать, что *теперь* состояние спина пропорционально $|\beta\rangle$, а при ответе **НЕТ**, что *теперь* состояние спина имеет направление, противоположное β. Но ни в одном из этих случаев ответ ничего не говорит нам о направлении α *до* измерения, а лишь дает нам некоторую вероятностную информацию о направлении α.

С другой стороны, в самом направлении α, вокруг которого электрон «вращается как вокруг оси» до того, как произведено измерение, по-видимому, есть нечто полностью *объективное*[11]. Действительно, мы *могли бы* остановить свой выбор на измерении спина электрона в направлении α, и электрон должен быть приготовлен так, чтобы *достоверно* (т. е. с вероятностью 100%) дать ответ **ДА**, если мы случайно угадаем истинное направление спина! Каким-то образом «информация» о том, что электрон действительно должен дать именно такой ответ, хранится в спиновом состоянии электрона.

Мне кажется, что при обсуждении вопроса о физической реальности в квантовой механике мы должны проводить различие между тем, что «объективно», и тем, что «измеримо». Действительно, вектор состояния системы несомненно *не измерим* в том смысле, что на основе экспериментов, произведенных над системой, невозможно определить (с точностью до коэффициента пропорциональности), каким является это состояние. Но очевидно, что вектор состояния является (опять-таки с точностью до коэффициента пропорциональности) *объективным* свойством системы, и полностью характеризуется результатами измерений, которые *могут быть* произведены над системой. В случае одной частицы со спином 1/2, например, электрона, такая объективность не является бессмысленной, так как она сводится просто к утверждению о том, что существует *некое* направление, относительно которого спин электрона точно определен, даже если мы не знаем, каково это направление. (Однако, как мы увидим в дальнейшем, такое представление относительно «объективности» в случае более сложных систем выглядит намного более странным — даже для системы, состоящей всего лишь из двух частиц со спинами 1/2.)

Но должен ли спин электрона *вообще* находиться в каком-нибудь физически определенном состоянии, прежде чем он будет измерен? Во многих случаях он *не имеет* определенного состояния, так как не может рассматриваться как автономная квантовая система. Вместо этого квантовое состояние в общем случае следует рассматривать как описание электрона, неразрывно связанного с большим числом других частиц. Но в особых случаях электрон (по крайней мере, если речь идет о его спине) *можно* рассматривать сам по себе. Например, в случае, когда спин электрона был точно измерен в некотором (возможно, неизвестном) направлении, а затем электрон в течение некоторого времени оставался невозмущенным, то его спин (в полном соответствии со стандартной квантовой теорией) объективно будет *иметь* вполне определенное направление.

Копирование квантового состояния

Объективность, но неизмеримость спинового состояния электрона поясняет еще один важный факт: *невозможно скопировать квантовое состояние, оставив оригинальное*

[11] Эта объективность является характерной особенностью нашего подхода, если мы всерьез принимаем стандартный квантовомеханический формализм. При *нестандартном* подходе система могла бы в действительности заранее «знать» результат, выдаваемый в ответ на *любое* измерение. Это привело бы нас к *другой* и, очевидно, объективной картине физической реальности.

состояние в неприкосновенном виде! Предположим, что мы могли бы изготовить копию спинового состояния электрона $|\alpha\rangle$. Если бы нам удалось сделать это один раз, то мы могли бы сделать это еще раз, а затем повторить еще и еще. Результирующая система имела бы огромный угловой момент вполне определенного направления. Это направление (обозначим его α) могло бы быть установлено с помощью макроскопического измерения. Но тогда оказалась бы нарушенной принципиальная *неизмеримость* спинового состояния $|\alpha\rangle$.

Но если мы готовы разрушить исходное состояние, то скопировать квантовое состояние все же возможно. Допустим, что у нас есть электрон в некотором неизвестном спиновом состоянии $|\alpha\rangle$ и нейтрон в некотором другом спиновом состоянии $|\gamma\rangle$. Вполне законно произвести обмен этими состояниями так, чтобы спиновым состоянием нейтрона стало $|\alpha\rangle$, а спиновым состоянием электрона $|\gamma\rangle$. То, что мы не можем — это изготовить спиновое состояние $|\alpha\rangle$ в *двух экземплярах* (если только мы уже не *знаем*, каково состояние $|\alpha\rangle$ на самом деле)! (См. также Вуттерс, Цурек [1982].)

Вспомним рассмотренную в главе 1 «машину для телепортации» (с. 58). Ее работа было основана на принципиальной возможности собрать на удаленной от нас планете полную копию тела и головного мозга какого-нибудь человека. Интригующе интересно предположить, что человеческое сознание может зависеть от некоторых аспектов квантового состояния. Если это так, то квантовая теория запрещала бы нам изготовление копии этого «сознания» без разрушения состояния оригинала — и тем самым можно было бы разрешить «парадокс» телепортации. Возможность существенного влияния квантовых эффектов на функционирование головного мозга будет рассмотрена в двух заключительных главах.

Спин фотона

Рассмотрим теперь «спин» фотона и его связь со сферой Римана. Фотоны действительно *обладают* спином, но поскольку они всегда движутся со скоростью света, их спин нельзя рассматривать как вращение вокруг какой-то неподвижной точки; ось спина фотона всегда совпадает с направлением движения. Спин фотона называется *поляризацией*. Поляризация — это явление, на котором основано действие «поляроидных» солнцезащитных очков. Возьмите два фрагмента поляроида, наложите их один на другой и посмотрите сквозь них. В общем случае вы увидите, что через них проходит некоторое количество света. Держа один из фрагментов неподвижно, поворачивайте другой фрагмент. Количество света, проходящего сквозь поляроиды, будет изменяться. При одной ориентации, когда проходит максимальное количество света, второй поляроид практически ничего не вычитает из светового потока, проходящего сквозь первый поляроид. Но при ориентации, выбранной под прямым углом к первой, свет практически вообще не проходит сквозь поляроиды.

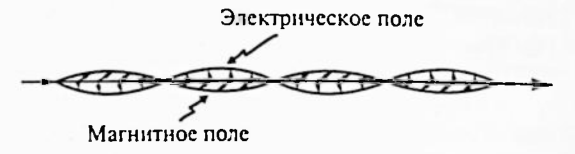
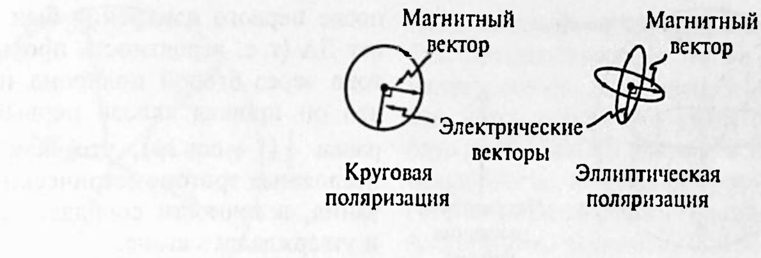


Рис. 6.26. Плоскополяризованная электромагнитная волна

Это явление легче всего понять в терминах волновой картины света. Здесь нам понадобится предложенный Максвеллом способ рассмотрения света как комбинации осциллирующих электрического и магнитного полей. На рис. 6.26 изображен *плоскополяризованный* свет. Электрическое поле осциллирует в плоскости, называемой *плоскостью поляризации*, а магнитное поле осциллирует в такт с электрическим, но в ортогональной плоскости. Каждый фрагмент поляроида пропускает свет, плоскость поляризации которого направлена вдоль структуры поляроида. Когда структура второго поляроида ориентирована так же, как структура первого, то весь свет, прошедший сквозь первый поляроид, проходит и сквозь второй. Но когда структуры двух поляроидов образуют прямой угол, то второй поляроид отсекает весь свет, прошедший сквозь первый поляроид. Если же два поляроида ориентированы друг относительно друга под некоторым углом φ, то второй поляроид



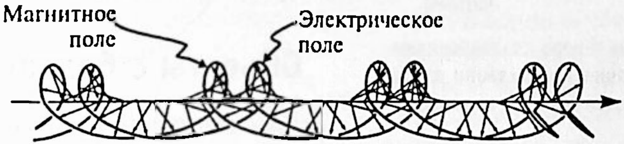


Рис. 6.27. Электромагнитная волна с круговой поляризацией. (Эллиптическая поляризация занимает промежуточное положение между плоской (рис. 6.26) и круговой (рис. 6.27) поляризацией.)

пропускает долю, равную

$$\cos^2\varphi,$$

света, прошедшего сквозь первый поляроид.

В корпускулярной картине мы должны считать, что *каждый индивидуальный фотон* обладает поляризацией. Первый поляроид действует как измеритель поляризации, давая ответ ДА, если фотон действительно поляризован в соответствующем направлении. В этом случае фотону разрешается пройти сквозь поляроид. Если же фотон поляризован в ортогональном направлении, то измерение первым поляроидом даст ответ НЕТ, и фотон будет поглощен. (В данном случае «ортогональность» в гильбертовом пространстве *соответствует прямому углу* между направлениями в обычном пространстве!) Предположим, что фотон проходит сквозь первый поляроид, после чего второй поляроид задает ему соответствующий вопрос, но уже относительно некоторого другого направления. Угол между этими двумя направлениями равен φ, как в упомянутом выше случае. Тогда мы имеем $\cos^2\varphi$ в качестве *вероятности* того, что фотон пройдет сквозь второй поляроид при условии, что он уже прошел сквозь первый поляроид.

Где же здесь появляется сфера Римана? Чтобы получить полный набор состояний поляризации, описываемый комплексными числами, нам необходимо рассмотреть *круговую* и *эллиптическую* поляризацию. Для классической волны эти разновидности поляризации представлены на рис. 6.27. При круговой поляризации электрическое и магнитное поля не осциллируют, а согласованно *вращаются*, по-прежнему образуя между собой прямой угол. При эллиптической поляризации существует некоторая комбинация вращательного и колебательного движений, а вектор электрического поля «вычерчивает» в пространстве *эллипс*. В квантовом описании каждому *индивидуальному* фотону разрешается находиться в любом из *спиновых* состояний, т. е. быть поляризованным любым из названных выше способов.

Чтобы понять, как набор возможных поляризаций снова образует сферу Римана, представим себе фотон, который движется вертикально вверх. Северный полюс теперь представляет состояние $|R\rangle$ — правовинтовой спин. Это означает, что электрический вектор движущегося фотона вращается против часовой стрелки относительно вертикали (если смотреть сверху). Южный полюс представляет состояние $|L\rangle$ — *левовинтовой* спин. (Фотоны можно представлять вращающимися наподобие ружейной пули, либо слева направо, либо справа налево.) Общее спиновое состояние $|R\rangle + q|L\rangle$ представляет собой комплексную линейную комбинацию двух состояний $|R\rangle$ и $|L\rangle$ и соответствует точке на сфере Римана, помеченной значением q. Чтобы установить связь между значением q и эллипсом поляризации, мы прежде всего извлечем из q *квадратный* корень и получим другое комплексное число p:

$$p = \sqrt{q}.$$

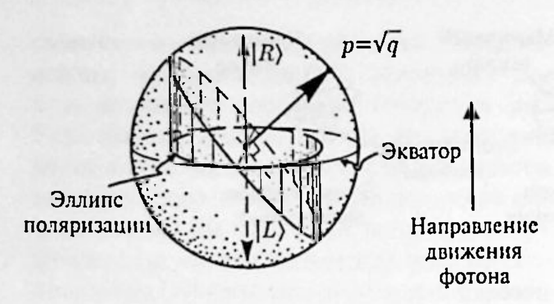


Рис. 6.28. Сфера Римана (но теперь со значениями $\sqrt{q}$) также описывает состояния поляризации фотона. (Вектор, направленный в точку $\sqrt{q}$, называется *вектором Стокса*.)

Затем нанесем p вместо q на сферу Римана и рассмотрим плоскость, проходящую через центр сферы перпендикулярно прямой, соединяющей центр сферы с точкой p. Эта плоскость пересекает сферу по окружности, проектируя которую на горизонталь, мы получаем эллипс поляризации (рис. 6.28) [12]. Сфера Римана со значениями q по-прежнему описывает совокупность поляризованных состояний фотона, но квадратный корень p из q дает нам ее пространственную реализацию.

Чтобы вычислить вероятности, мы можем воспользоваться той же самой формулой $\frac{1}{2}(1+\cos\vartheta)$, которой мы пользовались для электрона, применив ее к q, а не к p. Рассмотрим *плоскую* поляризацию. Мы измеряем поляризацию фотона сначала в одном направлении, затем в другом направлении, образующем с первым угол φ. Эти два направления соответствуют двум значениям p на экваторе сферы, стягивающим угол φ в центре сферы. Так как величины p — квадратные корни из величин q, угол ϑ, под которым из центра видны q-точки, *вдвое* больше угла, под которым из центра видны p-точки: $\vartheta = 2\varphi$. Таким образом, вероятность получения ответа ДА после второго измерения при условии, что после первого измерения был получен ответ ДА (т. е. вероятность прохождения фотона через второй поляроид при условии, что он прошел сквозь первый поляроид) равна $\frac{1}{2}(1+\cos 2\varphi)$, что, как показывают несложные тригонометрические преобразования, в точности совпадает с $\cos^2\varphi$, как и утверждалось выше.

Объекты с большим спином

Для квантовой системы с числом базисных состояний больше двух пространство физически различимых состояний имеет более сложную структуру, чем сфера Римана. Но в случае спина *самой* сфере Римана всегда отведена некоторая прямая геометрическая роль. Рассмотрим *массивную* частицу или атом со спином $n \times \hbar/2$ в состоянии покоя. (Для безмассовых частиц со спином, т. е. частиц, которые движутся со скоростью света (как, например, фотон), спин всегда, как было описано выше, представляет собой систему с *двумя* состояниями. Но у массивной частицы число состояний увеличивается с увеличением спина.) Если мы захотим измерить спин такой частицы в некотором направлении, то обнаружим, что существуют $n+1$ различных возможных исходов измерения, в зависимости от того, какая часть от полного спина ориентирована в выбранном направлении. В терминах фундаментальной единицы $\hbar/2$ возможные результаты для значений спина в выбранном направлении равны $n, n-2, n-4, \ldots, 2-n$ или $-n$. Следовательно, при $n=2$ спин может быть равен (в единицах $\hbar/2$) 2, 0 или -2, а при $n=3$ — 3, 1, -1 или -3 и т. д. *Отрицательные* значения соответствуют спину, направленному главным образом в сторону, *противоположную* той, в которой производилось измерение. В случае спина, равного 1/2, т. е. при $n=1$, значение 1 соответствует ответу ДА, а значение -1 — ответу НЕТ (в приведенных выше описаниях).

Оказывается, хотя я не буду пытаться излагать здесь причины (Майорана [1932], Пенроуз [1987a]), что *любое спиновое состояние* (с точностью до коэффициента пропорциональности) для спина $\hbar n/2$ однозначно характеризуется (неупорядоченным) *набором из n точек на* сфере Римана, т. е. n

[12] Комплексное число $-p$ подходит так же хорошо, как и p, в качестве квадратного корня из q, и дает тот же самый эллипс поляризации. Квадратный корень обусловлен тем, что фотон — безмассовая частица со *спином*, равным *единице*, т. е. *вдвое* бóльшим фундаментальной единицы $\hbar/2$. Для гравитона (еще не открытого кванта гравитации) спин равен *двум*, т. е. *вчетверо* бóльше фундаментальной единицы, поэтому нам в приведенном выше описании понадобился бы корень *четвертой* степени из q.

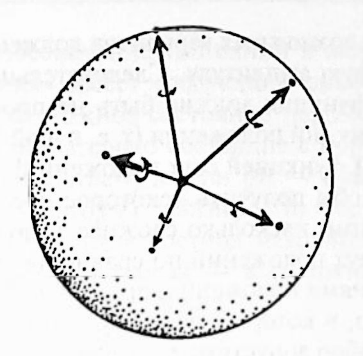


Рис. 6.29. Общее состояние с высшим спином для массивной частицы может быть описано как совокупность состояний со спином 1/2, ориентированных в произвольных направлениях

(обычно различными) направлениями из ее центра (рис. 6.29). (Эти направления определяются измерениями, которые могут быть произведены над системой: если мы измерим спин в одном из этих направлений, то результат заведомо не будет целиком ориентирован в противоположном направлении, т. е. даст одно из значений n, $n-2$, $n-4$, $\ldots$, $2-n$, но *не* $-n$.) В частном случае при $n=1$, как в приведенном выше примере с электроном, мы получим *одну* точку на сфере Римана. Это — просто точка, помеченная значением q в приведенных выше описаниях. Но для состояний с высшим спином картина, как я только что описал, значительно усложняется, хотя надо заметить, что это описание почему-то не очень знакомо физикам.

В этом описании есть нечто весьма удивительное. Часто высказывают мнение, что в некотором подходящем пределе квантовые описания атомов (или элементарных частиц, или молекул) с необходимостью переходят в классические ньютоновские описания, когда система увеличивается в размерах и усложняется. Но в такой формулировке *такое утверждение просто неверно*. Ибо, как мы только что видели, спиновые состояния объекта с большим угловым моментом соответствуют большому числу точек, разбросанных по сфере Римана[13]. Мы можем мысленно представлять себе спин объекта как состоящим из целого множества спинов 1/2, ориентированных по всем различным направлениям, задаваемыми этими точками. Лишь весьма немногие из таких комбинированных состояний, а именно когда большинство точек концентрируются вместе в небольшой области на сфере (т. е. когда большинство спинов 1/2 направлены примерно в одном и том же направлении), соответствуют реальным состояниям углового момента, которые мы обычно обнаруживаем у классических объектов, например, у крикетных шаров. Мы могли бы ожидать, что если выбрать спиновое состояние, в котором полный спин окажется равным (в единицах $\hbar/2$) некоторому очень большому числу, а в остальном это выбор будет «случайным», то начнет возникать нечто похожее на классический спин. Но в действительности все происходит совсем не так. В общем случае квантовые спиновые состояния с большим полным спином совсем не похожи на классические спиновые состояния!

Как же в таком случае следует устанавливать соответствие с угловым моментом из классической физики? Хотя большинство квантовых состояний с большим спином *не похожи* на классические состояния, они представляют собой линейные комбинации (ортогональных) состояний, каждое из которых *похоже* на классическое состояние. Каким-то образом над системой оказывается произведенным «измерение», и состояние «скачком» переходит в то или другое состояние, похожее на классическое. Ситуация здесь аналогична той, которая складывается с любым другим классически измеримым свойством системы, а не только с угловым моментом. Именно этот аспект квантовой механики должен вступать в игру всякий раз, когда система «выходит на классический уровень». Более подробно я расскажу об этом в дальнейшем, но прежде чем мы сможем обсудить такие «большие» или «сложные» квантовые системы, нам необходимо хотя бы несколько разобраться в том странном способе, которым квантовая механика пользуется при рассмотрении систем, состоящих более чем из одной частицы.

---

[13] Точнее, угловой момент описывается комплексными линейными комбинациями таких наборов из различного числа точек, так как суперпозиции могут включать несколько различных значений полного спинов — в случае какой-нибудь сложной системы. Все это приводит к картине, еще *менее* похожей на картину классического углового момента!

Многочастичные системы

Квантовомеханические описания многочастичных состояний, к сожалению, очень сложны. В действительности такие описания *чрезвычайно* сложны. О них необходимо думать в терминах суперпозиций *всех* различных возможных расположений всех отдельных частиц! Это приводит к огромному числу возможных состояний — гораздо большему, чем в случае *поля* в классической теории. Мы уже видели, что квантовое состояние даже *одной* частицы, а именно волновая функция, обладает сложностями такого рода, которые характерны для всего классического поля. Эта картина (требующая для своего задания *бесконечно* большого числа параметров) гораздо сложнее, чем классическая картина одной частицы (для задания состояния которой требуется всего лишь небольшое число параметров — точнее, шесть параметров, если частица не обладает внутренними степенями свободы, например, спином; см. главу 5, с. 168). Такая ситуация может показаться достаточно плохой, и можно было бы думать, что для описания квантового состояния двух частиц понадобится два *поля*, каждое из которых описывало бы состояние каждой частицы. Ничего подобного! Как мы увидим далее, в случае двух и более частиц описание квантового состояния становится гораздо сложнее.

Квантовое состояние одной (бесспиновой) частицы определяется комплексным числом (амплитудой) для каждого возможного положения, которое может занимать частица. Частица обладает амплитудой, чтобы находиться в точке A, и амплитудой, чтобы находиться в точке B, и амплитудой, чтобы находиться в точке C, и т. д. Подумаем теперь о *двух* частицах. Первая частица может находиться в точке A, а вторая, например, — в точке B. Возможность такого события должна была бы иметь некоторую амплитуду. С другой стороны, первая частица могла бы находиться в точке B, а вторая — в точке A, и такое расположение частиц также должно иметь некоторую амплитуду; возможно, что первая частица могла бы находиться в точке B, а вторая — в точке C или, может быть, обе частицы могли бы находиться в точке A. Каждый из этих возможных вариантов должен иметь некоторую амплитуду. Следовательно, волновая функция должна быть не просто парой функций положения (т. е. парой полей), а одной функцией *двух* положений!

Чтобы получить некоторое представление о том, насколько сложнее задать функцию двух положений по сравнению с двумя функциями положения, представим себе ситуацию, в которой существует лишь конечный набор допустимых положений. Предположим, что разрешены ровно 10 положений, заданных (ортонормированными) состояниями

$$|0\rangle, |1\rangle, |2\rangle, |3\rangle, |4\rangle, |5\rangle, |6\rangle, |7\rangle, |8\rangle, |9\rangle.$$

Тогда состояние $|\psi\rangle$ одной частицы было бы какой-то линейной комбинацией

$$|\psi\rangle = z_0|0\rangle + z_1|1\rangle + z_2|2\rangle + z_3|3\rangle + \ldots + z_9|9\rangle,$$

где различные коэффициенты $z_0, z_1, z_2, \ldots, z_9$ дают, соответственно, амплитуды того, что частица находится попеременно в каждой из 10 точек. Десять комплексных чисел задают состояние одной частицы. В случае *двухчастичного* состояния нам понадобилось бы по одной амплитуде для каждой *пары* положений. Всего существуют

$$10^2 = 100$$

различных (упорядоченных) пар положений, поэтому нам потребовались бы 100 комплексных чисел! А если бы у нас были только два одночастичных состояния (т. е. «две функции положения», а не «одна функция двух положений», как в приведенном выше примере), то нам понадобилось бы всего лишь 20 комплексных чисел.

Пронумеруем эти 100 комплексных чисел следующим образом

$$z_{00}, z_{01}, z_{02}, \ldots, z_{09}, z_{10}, z_{11}, \\ z_{12}, \ldots, z_{20}, \ldots, z_{99},$$

а соответствующие (ортонормированные) базисные векторы [12]

$$|0\rangle|0\rangle,\ |0\rangle|1\rangle,\ |0\rangle|2\rangle,\ \ldots \\ \ldots,\ |0\rangle|9\rangle, |1\rangle|0\rangle,\ \ldots,\ |9\rangle|9\rangle.$$

Тогда общее двухчастичное состояние можно было бы представить в виде

$$|\psi\rangle = z_{00}|0\rangle|0\rangle + z_{01}|0\rangle|1\rangle + \ldots + z_{99}|9\rangle|9\rangle.$$

Такое обозначение состояний в виде «произведения» имеет следующий смысл: если $|\alpha\rangle$ — возможное состояние первой частицы (не обязательно состояние с определенным положением) и если $|\beta\rangle$ — возможное состояние второй частицы, то состояние, в котором первая частица находится в состоянии $|\alpha\rangle$, а вторая — в состоянии $|\beta\rangle$, можно представить в виде

$$|\alpha\rangle|\beta\rangle.$$

«Произведения» можно также брать между любыми другими парами квантовых состояний, а не обязательно между парами одночастичных состояний. Таким образом, мы всегда интерпретируем состояние-произведение $|\alpha\rangle|\beta\rangle$ (не обязательно состояний отдельных частиц) как конъюнкцию

«первая система находится в состоянии $|\alpha\rangle$»
и
«вторая система находится в состоянии $|\beta\rangle$».

(Аналогичная интерпретация справедлива и относительно $|\alpha\rangle|\beta\rangle|\gamma\rangle$ и т. д.; см. далее.) Однако *общее* двухчастичное состояние в действительности не имеет вид «произведения». Например, оно может быть представимо в виде

$$|\alpha\rangle|\beta\rangle + |\rho\rangle|\sigma\rangle,$$

где $|\rho\rangle$ — еще одно возможное состояние первой системы, а $|\sigma\rangle$ — еще одно возможное состояние второй системы. Это состояние представляет собой *линейную суперпозицию*, а именно: суперпозицию первой конъюнкции состояний $|\alpha\rangle$ *и* $|\beta\rangle$ *плюс* вторая конъюнкция состояний $|\rho\rangle$ *и* $|\sigma\rangle$, и не может быть представлено в виде простого произведения (т. е. как конъюнкция двух состояний). Еще один пример — состояние $|\alpha\rangle|\beta\rangle - i|\rho\rangle|\sigma\rangle$ описывало бы другую такую линейную суперпозицию. Заметим, что квантовая механика требует проведения четкого различия между смыслом слов «плюс» и «и». И в обращении с этими словами нам следует быть более осторожными!

В случае трех частиц ситуация во многом аналогична. Чтобы задать общее трехчастичное состояние в приведенном выше примере, где имеются только 10 возможных положений, нам потребовалось бы теперь 1000 комплексных чисел! Полный базис для трехчастичных состояний состоял бы из следующих элементов:

$$|0\rangle|0\rangle|0\rangle,\ |0\rangle|0\rangle|1\rangle,\ |0\rangle|0\rangle|2\rangle,\ \ldots,\ |9\rangle|9\rangle|9\rangle.$$

Частные трехчастичные состояния имели бы вид произведений трех сомножителей

$$|\alpha\rangle|\beta\rangle|\gamma\rangle$$

(где $|\alpha\rangle$, $|\beta\rangle$ и $|\gamma\rangle$ — не обязательно состояния с определенным положением), но для общего трехчастичного состояния нам понадобилось бы построить суперпозицию большого числа состояний типа этих простых «произведений». Соответствующая схема получения общего состояния для четырех и более частиц должна быть очевидна.

До сих пор мы рассматривали случай *различимых* частиц, когда все частицы: «первая», «вторая», «третья» и т. д. принадлежат к *разным* типам. Одна из поразительных особенностей квантовой механики заключается в том, что в случае «тождественных» частиц правила коренным образом меняются. Действительно, правила становятся такими, что в самом прямом смысле частицы определенного типа должны быть не просто почти тождественными, а *в точности* тождественными. Это относится ко всем электронам и ко всем фотонам. Но оказывается, что все электроны тождественны друг другу *совсем не так*, как тождественны все фотоны! Различие заключается в том, что электроны принадлежат к так называемым фермионам, тогда как фотоны принадлежат к бозонам. Эти два класса частиц надлежит рассматривать весьма различным образом.

Прежде чем я окончательно запутаю читателя этими словесными несуразностями, позвольте мне попытаться объяснить, как действительно следует характеризовать фермионные и бозонные состояния. Правило состоит в следующем. Если $|\psi\rangle$ — состояние, содержащее некоторое число фермионов определенного типа, то при перестановке любых двух фермионов $|\psi\rangle$ должно перейти в $-|\psi\rangle$:

$$|\psi\rangle \to -|\psi\rangle.$$

Если состояние $|\psi\rangle$ содержит некоторое число бозонов определенного типа, то при перестановке любых двух бозонов $|\psi\rangle$ должно перейти в $|\psi\rangle$:

$$|\psi\rangle \to |\psi\rangle.$$

Отсюда следует, что *никакие два фермиона не могут находиться в одном и том же состоянии*. Действительно, если бы какие-нибудь два фермиона находились в одном и том же состоянии, то их перестановка вообще никак не сказывалась бы на полном состоянии системы, следовательно должно было бы выполняться $-|\psi\rangle = |\psi\rangle$, т. е. $|\psi\rangle = 0$, что не допустимо для квантового состояния. Это свойство известно как *принцип запрета Паули*[13], а его следствия для структуры вещества имеют фундаментальный характер. Действительно, все главные составляющие вещества: электроны, протоны и нейтроны принадлежат к числу фермионов. Не будь принципа запрета, вещество бы просто сколлапсировало!

Вернемся к нашему примеру с 10 положениями и предположим теперь, что у нас есть состояние, состоящее из двух тождественных фермионов. Состояние $|0\rangle|0\rangle$ исключается в силу принципа Паули (при перестановке первого множителя со вторым оно переходит в себя вместо того, чтобы переходить в себя со знаком минус). Кроме того, состояние $|0\rangle|1\rangle$ также само по себе должно быть исключено, так как при перестановке множителей знак минус не появляется; но это легко можно исправить, если заменить произведение $|0\rangle|1\rangle$ комбинацией

$$|0\rangle|1\rangle - |1\rangle|0\rangle.$$

(Для нормировки оба члена можно было бы умножить на общий множитель $1/\sqrt{2}$.) Это состояние правильно изменяет знак при перестановке первой частицы со второй, но теперь состояния $|0\rangle|1\rangle$ и $|1\rangle|0\rangle$ уже не независимы. Вместо этих *двух* состояний нам теперь разрешается иметь только *одно* состояние! Всего существует

$$\frac{1}{2}(10 \times 9) = 45$$

состояний такого рода — по одному на каждую неупорядоченную пару различных состояний из $|0\rangle, |1\rangle, \ldots, |9\rangle$. Таким образом, для задания двухфермионного состояния в нашей системе необходимы 45 комплексных чисел. В случае трех фермионов нам требуются 3 различные позиции, и базисные состояния выглядят следующим образом

$$|0\rangle|1\rangle|2\rangle + |1\rangle|2\rangle|0\rangle + |2\rangle|0\rangle|1\rangle - \\ - |0\rangle|2\rangle|1\rangle - |2\rangle|1\rangle|0\rangle - |1\rangle|0\rangle|2\rangle.$$

Всего таких состояний $(10 \times 9 \times 8)/6 = 120$, поэтому для задания трехфермионного состояния необходимы 120 комплексных чисел.

Для пары тождественных бозонов независимые базисные состояния бывают двоякого рода, а именно такие, как

$$|0\rangle|1\rangle + |1\rangle|0\rangle,$$

и такие, как

$$|0\rangle|0\rangle$$

(которое теперь разрешается), что дает всего $10 \times 11/2 = 55$ базисных состояний. Таким образом, для задания двухбозонных состояний требуется 55 комплексных чисел. Для трех бозонов существуют базисные состояния трех различных типов и для задания каждого из них требуются $(10 \times 11 \times 12)/6 = 220$ комплексных чисел, и так далее.

Разумеется, для того, чтобы донести до читателя основные идеи, я рассматривал упрощенную ситуацию. Более реалистическое описание потребовало бы целый континуум состояний с определенным положением, но существенные идеи остаются такими же. Еще одно небольшое осложнение связано с наличием *спина*. Для каждой частицы со спином 1/2 (такая частица с необходимостью является фермионом) в каждом положении существовало бы 2 возможных состояния. Обозначим их "↑" (спин «вверх») и "↓" (спин «вниз»). Тогда в рассматриваемой нами упрощенной ситуации мы получаем не 10, а 20 базисных состояний

$$|0\uparrow\rangle, |0\downarrow\rangle, |1\uparrow\rangle, |1\downarrow\rangle, |2\uparrow\rangle, |2\downarrow\rangle, \ldots, \\ \ldots, |9\uparrow\rangle, |9\downarrow\rangle,$$

а в остальном рассуждать следует так же, как было сделано только что (таким образом, для двух таких фермионов необходимо взять $(20 \times 19)/2 = 190$ чисел, для трех — $(20 \times 19 \times 18)/6 = 1140$ и т. д.).

В главе 1 я упоминал о том, что согласно современной теории, если частицу из тела человека поменять местами с аналогичной частицей из кирпича в стене его жилища, то ничего не произойдет. Если бы эта частица была бозоном, то, как мы знаем, состояние $|\psi\rangle$ действительно осталось бы совершенно не изменившимся. Если бы эта частица была фермионом, то состояние $|\psi\rangle$ в результате

обмена частиц перешло бы в $-|\psi\rangle$, физически тождественное состоянию $|\psi\rangle$. (В случае необходимости изменение знака можно устранить с помощью простой меры предосторожности, а именно: при замене одной частицы на другую, повернуть одну из двух частиц на 360° вокруг ее оси. Напомним, что фермионы изменяют знак при таком повороте, а состояние бозонов остается неизменным!) Современная теория (существующая примерно с 1926 года) действительно сообщает нам нечто глубокое относительно индивидуального тождества мельчайших «кирпичиков» физической материи. Строго говоря, мы не можем говорить об «этом конкретном электроне» или об «индивидуальном фотоне». Утверждать, что «первый электрон находится здесь, а второй — там», означает утверждать, что состояние имеет вид $|0\rangle|1\rangle$, что, как мы уже знаем, недопустимо, если речь идет о фермионном состоянии! Однако вполне допустимо утверждение о том, что «существует пара электронов, один из которых находится здесь, а другой — там». Вполне «законно» говорить о множестве всех электронов или всех протонов, или всех фотонов (хотя даже такое утверждение игнорирует *взаимодействия* между различными типами частиц). Индивидуальные электроны являются неким приближением такой полной картине, как, впрочем, и индивидуальные протоны или индивидуальные фотоны. Для большинства целей этого приближения вполне достаточно, но существуют различные ситуации, при которых оно не срабатывает — убедительными контрпримерами могут служить сверхпроводимость, сверхтекучесть и излучение лазера.

Картина физического мира, которую представила нам квантовая механика, — совсем не то, к чему мы привыкли в классической физике. Но придержите вашу шляпу — в квантовом мире есть гораздо более странные вещи!

«Парадокс» Эйнштейна, Подольского и Розена

Как упоминалось в начале этой главы, некоторые из идей Альберта Эйнштейна сыграли фундаментальную роль в развитии квантовой теории. Напомним, что именно Эйнштейн впервые ввел еще в 1905 году понятие «фотон» — квант электромагнитного поля — из этого понятия впоследствии выросла идея дуализма волна-частица. (Эйнштейну отчасти принадлежит и понятие «бозон», как и многие другие идеи, сыгравшие центральную роль в квантовой теории поля.) Тем не менее Эйнштейн так и не смог принять теорию, в которую впоследствии развились эти идеи, полагая, что такая теория не может быть описанием физического мира. Хорошо известно отвращение, которое Эйнштейн питал к вероятностному аспекту квантовой теории, и которое он в сжатой форме сформулировал в одном из писем к Максу Борну в 1926 году (письмо цитируется в книге: Пайс [1982], с. 443):

> «Квантовая механика производит очень внушительное впечатление. Но внутренний голос говорит мне, что это еще не настоящая „вещь“. Квантовая теория дает очень многое, но вряд ли способна приблизить нас к разгадке секрета Старика. Я глубоко убежден, что Он не играет в кости».

Однако, как оказывается, еще больше, чем такой физический индетерминизм, Эйнштейна беспокоило кажущееся *отсутствие объективности* в том, каким образом должна описываться квантовая теория. В моем изложении квантовой теории я пытался подчеркнуть, что описание мира, даваемое этой теорией, в действительности вполне объективно, хотя и кажется часто весьма странным и противоречащим интуиции. С другой стороны, Бор, по-видимому, считал, что квантовое состояние системы (между измерениями) не обладает настоящей физической реальностью, а действует лишь как свод «знаний некоторого субъекта» о рассматриваемой системе. Но разве различные наблюдатели не могут обладать различными знаниями о системе, тогда волновая функция должна была бы быть чем-то существенно *субъективным*, или «целиком существовать в уме физика»? Наша замечательно точная физическая картина мира, создававшаяся на протяжении многих столетий, не должна испариться целиком; поэтому Бору пришлось рассматривать мир на *классическом уровне* как действительно обладающий объективной реальностью.

Но в состояниях на *квантовом* уровне, которые, казалось бы, лежат в основе всего, никакой «реальности» он не усматривал.

Такая картина была неприемлема для Эйнштейна, который был глубоко убежден в том, что объективный физический мир должен действительно существовать, даже на микроскопических масштабах квантовых явлений. В своих многочисленных дискуссиях с Бором Эйнштейн пытался (но неудачно) показать, что квантовой картине присущи внутренние противоречия, и что за квантовой теорией должна стоять какая-то более глубокая структура, возможно, более похожая на картины классической физики. Возможно, вероятностное поведение квантовых систем является проявлением статистических эффектов более малых компонентов, или частей, системы, о которых мы не располагаем непосредственным знанием. Последователи Эйнштейна, в особенности Давид Бом, развили высказанную им идею о «скрытых переменных», согласно которой должна существовать некоторая вполне определенная реальность, но параметры, точно определяющие систему, не доступны нам непосредственно, и квантовые вероятности возникают из-за того, что значения этих параметров неизвестны до измерения.

Согласуется ли теория скрытых переменных со всеми наблюдаемыми фактами квантовой физики? Похоже, что ответ на этот вопрос должен быть утвердительным, но только если эта теория по существу *нелокальна* в том смысле, что скрытые параметры должны иметь возможность мгновенно влиять на элементы системы в сколь угодно далеких областях! Такая ситуация не понравилась бы Эйнштейну, особенно в связи с возникающими трудностями в специальной теории относительности. К ним я еще вернусь в дальнейшем. Наиболее успешная теория скрытых переменных известна как модель де Бройля (де Бройль [1956], Бом [1952]). Я не буду обсуждать здесь эти модели, так как в этой главе моя цель состоит только в том, чтобы дать общий обзор стандартной квантовой теории, а не различных соперничающих с ней положений. Если кто-нибудь жаждет физической реальности, но готов пожертвовать детерминизмом, то самой стандартной теории вполне достаточно. Он просто рассматривает вектор состояния как описывающий «реальность» — обычно изменяющийся во времени в соответствии с гладкой детерминистской $\mathbf{U}$-процедурой, но время от времени совершающий причудливые «прыжки» в соответствии с $\mathbf{R}$-процедурой всякий раз, когда эффект увеличивается до классического уровня. Но проблема нелокальности и явных трудностей с относительностью сохраняются. Рассмотрим некоторые из них.

Предположим, что у нас имеется физическая система, состоящая из двух подсистем A и B. Пусть, например, A и B — две различные частицы. Предположим, что для состояния частицы A существуют две (ортогональные) альтернативы $|\alpha\rangle$ и $|\rho\rangle$, а для состояния частицы B — две (ортогональные) альтернативы $|\beta\rangle$ и $|\sigma\rangle$. Как мы уже видели выше, общее комбинированное состояние системы будет не просто произведением (конъюнкцией «и») некоторого состояния частицы A и некоторого состояния частицы B, а суперпозицией («плюс») таких произведений. (Тогда мы говорим, что A и B *коррелированы*.) Пусть состояние системы представимо суперпозицией

$$|\alpha\rangle|\beta\rangle + |\rho\rangle|\sigma\rangle.$$

Произведем измерение типа «да или нет» над частицей A, которое отличает состояние $|\alpha\rangle$ (ДА) от состояния $|\rho\rangle$ (НЕТ). Что произойдет при этом с частицей B? Если измерение даст ответ ДА, то результирующим должно быть состояние

$$|\alpha\rangle|\beta\rangle,$$

а если измерение даст ответ НЕТ, то

$$|\rho\rangle|\sigma\rangle.$$

Таким образом, измерение, производимое нами над частицей A, заставляет состояние частицы B измениться скачком: перейти в $|\beta\rangle$, если получен ответ ДА, и перейти в $|\sigma\rangle$, если получен ответ НЕТ! Частица B не обязательно должна находиться поблизости от частицы A; частицы могут быть разделены расстоянием в несколько световых лет. И все же частица B скачком переходит из одного состояния в другое одновременно с измерением, производимым над частицей A!

«Но постойте», — вполне может сказать читатель. К чему все эти подозрительные «скачки»? Почему не происходит просто следующее: представьте себе ящик, о котором известно, что в нем лежит один черный и один белый шар. Предположим, что некто извлек шары из ящика и, не глядя, отнес их в противоположные углы комнаты. Затем он взглянул на один шар и обнаружил, что он белый (аналог упоминавшегося выше состояния $|\alpha\rangle$), тогда — алле-оп! —другой шар оказывается черным (аналог состояния $|\beta\rangle$)! С другой стороны, если первый шар оказался черным (аналог состояния $|\rho\rangle$), то в мгновение ока состояние второго шара скачком переходит в «заведомо белый» (аналог состояния $|\sigma\rangle$). Никто из читателей или читательниц в здравом уме не станет упорно приписывать внезапный переход второго шара из состояния «неопределенности» в состояние «определенно черный» или «определенно белый» некоторому таинственному нелокальному «влиянию», мгновенно доходящему до него от первого шара в тот самый момент, когда наблюдатель рассмотрел первый шар.

Но природа действует еще более изощренно. Действительно, в приведенном выше примере можно было бы представить, что *система* уже «знала», что частица B находилась в состоянии $|\beta\rangle$, а частица A — в состоянии $|\alpha\rangle$ (или что частица B находилась в состоянии $|\sigma\rangle$, а частица A — в состоянии $|\rho\rangle$) до того, как над A было произведено измерение; и только *экспериментатору* состояния частиц не были известны. Обнаружив, что частица A находится в состоянии $|\alpha\rangle$, он просто *заключил*, что частица B находится в состоянии $|\beta\rangle$. Такая точка зрения была бы «классической» — как в локальной теории скрытых переменных — и никаких скачкообразных *физических* переходов из одного состояния в другое в действительности не происходит. (Все это происходит лишь в уме экспериментатора!) Согласно такой точке зрения любая часть системы заранее «знает» результаты любого эксперимента, который мог бы быть произведен над ней. Вероятности возникают только из-за отсутствия такого знания у экспериментатора. Достойно удивления, что, как оказывается, эта точка зрения *не срабатывает* для объяснения всех загадочных нелокальных вероятностей, возникающих в квантовой теории!

Чтобы убедиться в этом, рассмотрим ситуацию, аналогичную изложенной выше, но такую, что *выбор измерения*, производимого над системой A, остается нерешенным до тех пор, пока системы A и B не окажутся пространственно разделенными. Тогда, как представляется, факт выбора измерения мгновенно окажет влияние на поведение системы B! Этот кажущийся парадоксальным «мысленный эксперимент» (ЭПР-типа) был предложен Альбертом Эйнштейном, Борисом Подольским и Натаном Розеном [1935]. Я опишу его вариант, предложенный Давидом Бомом [1951]. То, что никакое локальное «реалистическое» (т. е. типа скрытых переменных или «классического типа») описание не может дать правильные квантовые вероятности, следует из одной замечательной теоремы Джона С. Белла (Белл [1987], Рэй [1986], Сквайерс [1986]).

Предположим, что две частицы со спином 1/2, которые я буду называть *электроном* и *позитроном* (т. е. *антиэлектроном*), возникли в результате распада одной частицы со спином 0 в некоторой точке (центре), и что они движутся от центра в противоположных направлениях (рис. 6.30). Из закона сохранения углового момента следует, что спины электрона и позитрона в сумме должны давать 0, так как угловой момент исходной частицы был равен 0. Отсюда следует, что когда мы измеряем спин электрона в каком-нибудь направлении, то, какое

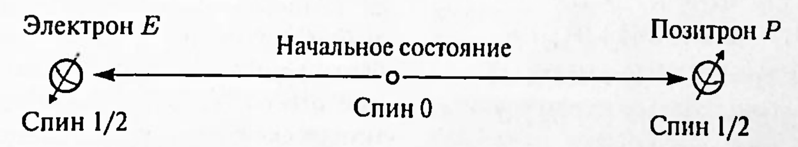


Рис. 6.30. Частица с нулевым спином распадается на две частицы с половинным спином — электрон E и позитрон P. Представляется, что измерение спина одной из частиц со спином 1/2 *мгновенно* фиксирует состояние спина другой частицы

направление мы бы ни выбрали, спин позитрона окажется направленным в *противоположную* сторону! Электрон и позитрон могут быть разделены расстоянием в несколько миль или даже световых лет, тем не менее кажется, что сам *выбор* измерения, производимого над одной частицей, *мгновенно* фиксирует ось спина другой частицы!

Попытаемся теперь выяснить, как квантовый формализм приводит нас к такому заключению. Представим состояние двух частиц с суммарным нулевым угловым моментом вектором состояния $|Q\rangle$. Тогда имеем соотношение

$$|Q\rangle = |E\uparrow\rangle|P\downarrow\rangle - |E\downarrow\rangle|P\uparrow\rangle,$$

где E означает электрон, а P — позитрон. Здесь все описывается в терминах направлений спина «вверх/вниз». Мы видим, что полное состояние является линейной суперпозицией электрона со спином вверх и позитрона со спином вниз, а также электрона со спином вниз и позитрона со спином вверх. Таким образом, если мы измеряем спин электрона в направлении «вверх/вниз» и обнаруживаем, что спин направлен вверх, то мы должны скачком перейти к состоянию $|E\uparrow\rangle|P\downarrow\rangle$, поэтому спиновое состояние позитрона должно быть направлено вниз. С другой стороны, если мы обнаруживаем, что спин электрона направлен вниз, то состояние скачком переходит в $|E\downarrow\rangle|P\uparrow\rangle$, поэтому спин позитрона направлен вверх.

Предположим, что мы выбрали какую-то другую пару противоположных направлений, например, вправо и влево, где

$$|E\rightarrow\rangle = |E\uparrow\rangle + |E\downarrow\rangle, \quad |P\rightarrow\rangle = |P\uparrow\rangle + |P\downarrow\rangle$$

и

$$|E\leftarrow\rangle = |E\uparrow\rangle - |E\downarrow\rangle, \quad |P\leftarrow\rangle = |P\uparrow\rangle - |P\downarrow\rangle.$$

Тогда мы находим (если угодно, можете проверить выкладки):

$$\begin{aligned}
&|E\rightarrow\rangle|P\leftarrow\rangle - |E\leftarrow\rangle|P\rightarrow\rangle = \\
&= (|E\uparrow\rangle + |E\downarrow\rangle)(|P\uparrow\rangle - |P\downarrow\rangle) - \\
&\quad - (|E\uparrow\rangle - |E\downarrow\rangle)(|P\uparrow\rangle + |P\downarrow\rangle) = \\
&= |E\uparrow\rangle|P\uparrow\rangle + |E\downarrow\rangle|P\uparrow\rangle - |E\uparrow\rangle|P\downarrow\rangle - \\
&\quad - |E\downarrow\rangle|P\downarrow\rangle - |E\uparrow\rangle|P\uparrow\rangle + |E\downarrow\rangle|P\uparrow\rangle - \\
&\quad - |E\uparrow\rangle|P\downarrow\rangle + |E\downarrow\rangle|P\downarrow\rangle = \\
&= -2(|E\uparrow\rangle|P\downarrow\rangle - |E\downarrow\rangle|P\uparrow\rangle) = \\
&= -2|Q\rangle,
\end{aligned}$$

т. е. мы получили (с точностью до несущественного множителя -2) то же самое состояние, из которого мы «стартовали». Таким образом, наше исходное состояние можно одинаково хорошо считать линейной суперпозицией электрона со спином вправо, позитрона со спином влево, и электрона со спином влево, позитрона со спином вправо! Выписанное выше выражение полезно, если мы решили измерять спин электрона в направлении вправо-влево вместо направления вверх-вниз. Если мы обнаружим, что спин электрона действительно направлен вправо, то состояние системы скачком переходит в $|E\rightarrow\rangle|P\leftarrow\rangle$, поэтому спин позитрона направлен влево. С другой стороны, если мы обнаружим, что спин электрона направлен влево, то состояние системы скачком переходит в $|E\leftarrow\rangle|P\rightarrow\rangle$, поэтому спин позитрона направлен вправо. Если бы мы стали измерять спин электрона в любом другом направлении, то получили бы соответствующую ситуацию: спиновое состояние позитрона мгновенно перешло бы скачком либо в измеряемое направление, либо в противоположное направление, в зависимости от измерения спина электрона.

Почему мы не можем моделировать спины наших частиц — электрона и позитрона аналогично тому, как мы поступили в приведенном выше примере с черным и белым шарами, извлекаемыми из ящика? Будем рассуждать на самом общем уровне. Вместо черного и белого шаров мы могли бы взять два каких-нибудь технических устройства E и P, первоначально образовывавших единое целое, а затем начавших двигаться в противоположные стороны. Предположим, что каждое из устройств E и P способно давать ответ ДА или НЕТ на измерение спина в любом заданном направлении. Этот ответ может полностью определяться технической начинкой устройства при любом выборе направления — или, может быть, устройство дает только вероятностные ответы (вероятность определяется его технической начинкой) — но при этом мы предполагаем, что после разделения *каждое из устройств E и P ведет себя совершенно независимо от другого.*

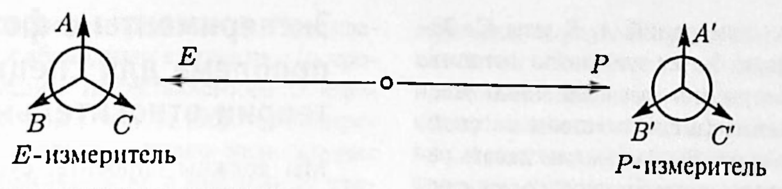


Рис. 6.31. Простая версия парадокса ЭПР, принадлежащая Дэвиду Мермину, и теорема Белла, показывающие, что существует противоречие между локальным реалистическим взглядом на природу и результатами квантовой теории. E-измеритель и P-измеритель каждый независимо имеет по три настройки для направлений, в которых они могут измерять спины соответствующих частиц (электрона и позитрона)

Поставим с каждой стороны измерители спина, один из которых измеряет спин E, а другой — спин P. Предположим, что каждый измеритель обладает тремя настройками для измерения направления спина при каждом измерении, например, настройками A, B, C для измерителя спина E и настройками A', B', C' для измерителя спина P. Направления A', B', C' должны быть параллельны, соответственно, направлениям A, B и C. Предполагается также, что все три направления A, B и C лежат в одной плоскости и образуют между собой попарно равные углы, т. е. углы в $120°$ (рис. 6.31). Предположим теперь, что эксперимент повторяется многократно и дает различные результаты для каждой из настроек. Иногда E-измеритель фиксирует ответ ДА (т. е. спин направлен вдоль измеряемого направления A, B или C), иногда фиксирует ответ НЕТ (т. е. спин имеет направление, противоположное тому, в котором производится измерение). Аналогично, P-измеритель фиксирует иногда ответ ДА, иногда — НЕТ. Обратим внимание на два свойства, которыми должны обладать настоящие *квантовые* вероятности:

(1) Если настройки устройств E и P *одинаковы* (т. е. A совпадает с A' и т. д.), то результаты измерений, производимых с помощью устройств E и P, всегда *не согласуются между собой* (т. е. E-измеритель фиксирует ответ ДА всякий раз, когда P-измеритель дает ответ НЕТ, и ответ НЕТ всякий раз, когда P-измеритель даст ответ ДА).

(2) Если лимбы настроек могут вращаться и установлены *случайно*, т. е. полностью независимо друг от друга, то два измерителя *равновероятно дают как согласующиеся, так и не согласующиеся* результаты измерений.

Нетрудно видеть, что свойства (1) и (2) непосредственно следуют из приведенных выше правил квантовых вероятностей. Мы можем предположить, что E-измеритель срабатывает первым. Тогда P-измеритель обнаруживает частицу, спиновое состояние которой имеет направление, противоположное измеренному E-измерителем, поэтому свойство (1) следует немедленно. Чтобы получить свойство (2), заметим, что для измеряемых направлений, образующих между собой углы в $120°$, если E-измеритель дает ответ ДА, то P-направление расположено под углом $60°$ к тому спиновому состоянию, на которое действует P-измеритель, а если E-измеритель дает ответ НЕТ, то P-направление образует угол $120°$ с этим спиновым состоянием. С вероятностью $3/4 = (1/2)(1 + \cos 60°)$ измерения согласуются, и с вероятностью $1/4 = (1/2)(1 + \cos 120°)$ они не согласуются. Таким образом, усредненная вероятность для трех настроек P-измерителя при условии, что E-измеритель дает ответ ДА, составляет $(1/3)(0 + 3/4 + 3/4) = 1/2$ для ответа ДА, даваемого P-измерителем, и $(1/3)(1 + 1/4 + 1/4) = 1/2$ для ответа НЕТ, даваемого P-измерителем, т. е. результаты измерений, производимых E- и P-измерителями, равновероятно согласуются и не согласуются. Аналогичная ситуация возникает и в том случае, когда E-измеритель дает ответ НЕТ. Это и есть свойство (2) (см. с. 238).

Замечательно, что свойства (1) и (2) *не согласуются* с любой локальной реалистической моделью (т. е. с любой разновидностью устройств рассматриваемого типа)! Предположим, что у нас есть такая модель. E-машину следует приготовить для каждого

из возможных измерений A, B или C. Заметим, что если бы ее следовало готовить только для получения *вероятностного* ответа, то P-машина (в соответствии со свойством (1)) не могла бы *достоверно* давать результаты измерения, не согласующиеся с результатами измерения E-машины. Действительно, *обе* машины должны давать свои ответы, определенным образом приготовленные заранее, на каждое из трех возможных измерений. Предположим, например, что эти ответы должны быть ДА, ДА, ДА, соответственно, для настроек A, B, C; тогда правая частица должна быть приготовлена так, чтобы давать ответы НЕТ, НЕТ, НЕТ при соответствующих трех настройках. Если же вместо этого приготовленные ответы левой частицы гласят: ДА, ДА, НЕТ, то ответами правой частицы должны быть НЕТ, НЕТ, ДА. Все остальные случаи по существу аналогичны только что приведенным. Попытаемся теперь выяснить, согласуется ли это со свойством (2). Наборы ответов ДА, ДА, ДА / НЕТ, НЕТ, НЕТ не слишком многообещающи, так как дают 9 случаев несоответствия и 0 случаев соответствия при всех возможных парах настроек A/A', A/B', A/C', B/A' и т.д. А как обстоит дело с наборами ДА, ДА, НЕТ / НЕТ, НЕТ, ДА и тому подобными ответами? Они дают 5 случаев несоответствия и 4 случая соответствия. (Чтобы убедиться в правильности последнего утверждения, произведем подсчет случаев: Д/Н, Д/Н, Д/Д, Д/Н, Д/Н, Д/Д, Н/Н, Н/Н, Н/Д. Мы видим, что в 5 случаях ответы не согласуются и в 4 случаях согласуются.) Это уже гораздо ближе к тому, что требуется для свойства (2), но еще *недостаточно* хорошо, так как случаев несоответствия ответов должно быть столько же, сколько случаев соответствия! Для любой другой пары наборов возможных ответов, согласующихся со свойством (1), мы снова получили бы соотношение 5 к 4 (за исключением наборов НЕТ, НЕТ, НЕТ / ДА, ДА, ДА, для которых соотношение было бы хуже — снова 9 к 0). *Не существует* набора приготовленных ответов, который могли бы дать квантово-механические вероятности. *Локальные реалистические модели исключаются!* [14]

Эксперименты с фотонами: проблема для специальной теории относительности?

Мы должны спросить, существуют ли реальные эксперименты, которые подкрепляют эти удивительные квантовые ожидания? Только что описанный точный эксперимент — гипотетический, он никогда не был осуществлен на самом деле. Но были осуществлены похожие эксперименты, в которых использовалась поляризация пары *фотонов*, а не спин массивных частиц со спином 1/2. Кроме этого различия проведенные эксперименты не отличались в принципе от описанного выше гипотетического эксперимента — за исключением того, что фигурировавшие в них углы были вдвое меньше углов для частиц со спином 1/2 (так как спин фотона равен 1, а не 1/2). Поляризации пар фотонов были измерены в нескольких различных комбинациях направлений, и результаты оказались в полном соответствии с предсказаниями квантовой теории, и не согласовывались ни с какой локальной реалистической моделью!

Наиболее точные и убедительные экспериментальные результаты, полученные к настоящему времени, принадлежат Алену Аспекту [1986] и его коллегам из Парижа [15]. Эксперименты Аспекта обладают еще одной интересной особенностью. «Выбор» способа измерения поляризаций фотонов определялся только после испускания фотонов, когда они уже находились в полете. Таким образом, если мы мысленно представим себе некоторое нелокальное «влияние», распространяющееся от детектора одного фотона к фотону, находящемуся на противоположной стороне, и сигнализирующее о направлении, в котором экспериментатор намеревается измерить направление поляризации приближающегося фотона, то придем к заключению, что это «влияние» должно распространяться быстрее света! Ясно, что любое реалистическое описание квантового мира, согласующееся с этими фактами, должно быть *не-причинным* в том смысле, что влияние должно обладать способностью распространяться быстрее света!

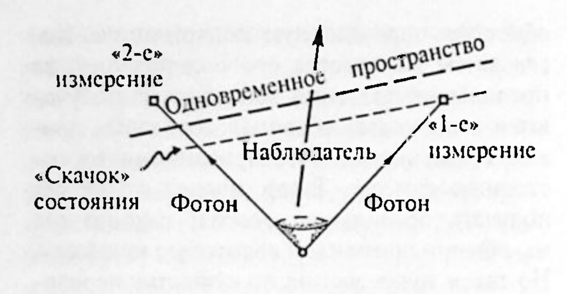
Но в предыдущей главе мы видели, что в силу теории относительности, испускание

сигналов, распространяющихся быстрее света, приводит к абсурдным ситуациям (и противоречит нашим представлениям о «свободе воли» и т. д.; см. с. 196). Это определенно справедливо, однако нелокальные «влияния», возникающие в мысленных экспериментах типа ЭПР, не таковы, чтобы их можно было использовать для отправления сообщений (по той самой причине, как это нетрудно понять, что это могло бы приводить к абсурдным ситуациям). (Подробное доказательство того, что такие «влияния» не могут быть использованы для испускания сигналов и передачи сообщений, было дано Гирарди, Римини и Вебером [1980].) Бесполезно знать, что фотон поляризован «либо вертикально, либо горизонтально» (или, наоборот, «либо под углом 60°, либо 150°») до тех пор, пока экспериментатор не информирован, какая из альтернатив соответствует действительности. Именно эта часть «информации» (т. е. альтернативные *направления* поляризации) распространяется быстрее света («мгновенно»), тогда как информация о том, в *каком* из двух направлений действительно поляризован фотон, доходит до экспериментатора медленнее и через обычный сигнал, сообщающий *результат* первого измерения поляризации.

Хотя эксперименты типа ЭПР не противоречат (в обычном смысле передачи сообщений сигналами) *причинности* специальной теории относительности, существует определенный конфликт с *духом* теории относительности в нашей картине физической реальности. Попытаемся выяснить, каким образом *реалистическая* точка зрения, основанная на использовании понятия вектора состояния, применима к описанному выше эксперименту типа ЭПР (с фотонами). Когда два фотона разлетаются, вектор состояния описывает *пару* фотонов, действующих как единое целое. Ни один из фотонов в отдельности не обладает объективным состоянием: квантовое состояние применимо только к двум фотонам вместе. Ни один из фотонов в отдельности не обладает направлением поляризации: поляризация — комбинированное свойство двух фотонов вместе. При измерении поляризации одного из этих фотонов вектор состояния изменяется *скачком*, так что неизмеряемый фотон *обретает* определенную поляризацию. Когда затем измеряется его поляризация, то правильные значения вероятности получаются с помощью обычных квантовых правил, применяемых к поляризационному состоянию фотона. Такой подход позволяет получать правильные ответы; именно так мы обычно применяем квантовую механику. Но такая точка зрения по существу нерелятивистская. Действительно, два измерения поляризации разделены *пространственноподобным интервалом*. Это означает, что каждое измерение лежит вне светового конуса другого, как точки R и Q на рис. 5.21. Вопрос о том, какое из этих измерений произведено *первым*, не имеет реального физического смысла, а зависит от состояния движения «наблюдателя» (рис. 6.32). Если «наблюдатель» достаточно быстро движется вправо, то измерение, производимое справа, он считает происходящим первым; а если «наблюдатель» движется влево, то первым он считает измерение, производимое слева. Но если мы сочтем, что первым был измерен правый фотон, то получим совершенно другую картину физической реальности, чем та, которая получается, если мы сочтем, что первым был измерен левый фотон! (Это — другое измерение, вызывающее нелокальный «скачок».) Между нашей пространственно-временной картиной физической реальности (даже правильной нелокальной квантово-механической картиной) и специальной теорией относительности имеется существенное противоречие! Это — трудная задача, адекватное решение которой не удалось пока решить «квантовым реалистам» (см. Ааронов, Альберт [1981]). К этому вопросу мне еще придется вернуться в дальнейшем.

Уравнение Шредингера; уравнение Дирака

Выше в этой главе я уже упоминал об уравнении Шредингера, которое является хорошо определенным детерминистским уравнением, во многих отношениях аналогичным уравнениям классической физики. Правила гласят, что до тех пор, пока над квантовой системой не производятся «измерения» (или «наблюдения»), уравнение



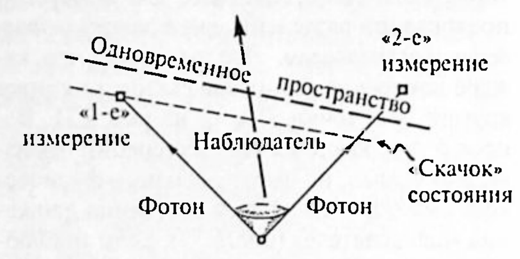


Рис. 6.32. У двух различных наблюдателей формируются взаимно несогласованные картины «реальности» в эксперименте ЭПР, в котором два фотона в состоянии со спином 0 испускаются в противоположных направлениях. С точки зрения наблюдателя, движущегося вправо, левая часть состояния совершает скачок *до* того, как производится измерение, где скачок обусловлен измерением, производимым над правой частью состояния. Наблюдатель, движущийся влево, придерживается противоположного мнения!

Шредингера должно оставаться справедливым. Читатель может захотеть узнать, как выглядит уравнение Шредингера в явном виде:

$$i\hbar\frac{\partial}{\partial t}|\psi\rangle = H|\psi\rangle.$$

Напомним, что $\hbar$ — дираковский вариант постоянной Планка ($h/2\pi$) (мнимая единица $i = \sqrt{-1}$), оператор $\partial/\partial t$ (частного дифференцирования по времени), действующий на $|\psi\rangle$, просто означает *скорость* изменения состояния $|\psi\rangle$ со временем. Уравнение Шредингера означает, что эволюцию состояния $|\psi\rangle$ описывает величина $H|\psi\rangle$.

Но что такое "H"? Это — *функция Гамильтона*, которую мы рассматривали в предыдущей главе, но с одним принципиальным различием! Напомним, что классическая функция Гамильтона, или гамильтониан, — это выражение для *полной энергии* через различные координаты положения q_i и импульсные координаты p_i всех физических объектов, входящих в систему. Чтобы получить *квантовый* гамильтониан, мы берем то же самое выражение, но вместо каждого импульса p_i подставляем *дифференциальный оператор*, кратный оператору частного дифференцирования по q_i. В частности, p_i мы заменяем на $-i\hbar\partial/\partial q_i$. В результате наш квантовый гамильтониан H становится некоторой (нередко сложной) математической *операцией*, включающей в себя дифференцирование и умножение (причем не только на число!) и т. д. Это выглядит, как фокус-покус! Но дело не просто в исполнении математических трюков; в действительности перед нами самая настоящая *магия*! (Некая толика «искусства» заключена уже в самом процессе получения квантового гамильтониана из классического, но еще более удивительно, имея в виду его «экстравагантную» природу, что неоднозначности, присущие этой процедуре, не играют сколь-нибудь существенную роль.)

Относительно уравнения Шредингера (что бы ни означало H) важно заметить, что оно *линейное*, т. е. если $|\psi\rangle$ и $|\varphi\rangle$ оба удовлетворяют уравнению Шредингера, то ему также удовлетворяет $|\psi\rangle + |\varphi\rangle$, а в действительности любая комбинация $w|\psi\rangle + z|\varphi\rangle$, где w и z — заданные комплексные числа. Таким образом, комплексная линейная суперпозиция удовлетворяет уравнению Шредингера неограниченно долго. (Комплексная) линейная суперпозиция двух возможных альтернативных состояний не может быть «расщеплена» действием одного лишь оператора U! Именно поэтому необходимо действие оператора R как *отдельной* процедуры, чтобы в конце концов выжило всего лишь *одно* альтернативное состояние.

Подобно гамильтоновому формализму в классической физике, уравнение Шредингера не является лишь конкретным отдельным уравнением, а служит общей схемой для квантовомеханических уравнений. Если для решаемой задачи удалось получить квантовый гамильтониан, то эволюция состояния (его развитие во времени) в соответствии с уравнением Шредингера происходит так, как если бы $|\psi\rangle$ было каким-нибудь классическим полем, удовлетворяющим некоторому классическому полевому уравнению, например, уравнениям Максвелла. Действительно, если $|\psi\rangle$ описывает состояние отдельного *фотона*, то

оказывается, что уравнение Шредингера *переходит* в уравнения Максвелла! Уравнение для отдельного фотона есть в точности то самое уравнение [14], которое было выведено для всего электромагнитного поля. Именно этим обстоятельством обусловлено волнообразное поведение фотона, аналогичное поведению электромагнитного поля Максвелла, и поляризация *отдельных фотонов* — эффекты, с которыми мы бегло ознакомились ранее. В качестве еще одного примера упомянем о том, что если $|\psi\rangle$ описывает состояние одного *электрона*, то уравнение Шредингера переходит в замечательное волновое уравнение Дирака, открытое в 1928 году после того, как Дирак приложил к его выводу немало проницательности и оригинальных идей.

В действительности уравнение Дирака для электрона по праву должно считаться наряду с уравнениями Максвелла и Эйнштейна одним из великих полевых уравнений физики. Чтобы создать у читателя адекватное представление об уравнении Дирака, мне понадобилось бы ввести здесь математические понятия, которые не столько проясняли суть дела, сколько затемнили бы его еще больше. Достаточно сказать, что в уравнении Дирака $|\psi\rangle$ обладает любопытным «фермионным» свойством $|\psi\rangle \to -|\psi\rangle$ при повороте на 360°, о котором мы упоминали выше (с. 236). Уравнения Дирака и Максвелла являются фундаментальными составляющими квантовой электродинамики, самой успешной из всех квантовых теорий поля. Давайте ознакомимся вкратце с этой теорией.

Квантовая теория поля

Предмет, известный под названием «квантовая теория поля», возник из объединения идей специальной теории относительности и квантовой механики. От стандартной (т. е. нерелятивистской) квантовой механики квантовая теория поля отличается тем, что число частиц (любого рода) в ней не обязательно постоянно. Для каждого рода частицы существует ее *античастица* (иногда, как в случае фотонов, античастица и частица совпадают). Массивная частица и ее античастица могут аннигилировать с выделением энергии. С другой стороны, пара частица-античастица может рождаться из энергии. Действительно, число частиц не обязательно должно быть даже определенным, ибо допускаются линейные суперпозиции состояний с различным числом частиц. «Верховной» квантовой теорией поля по праву считается «квантовая электродинамика» — по сути, теория электронов и протонов. Квантовая теория поля замечательна точностью своих предсказаний (например, она предсказала точное значение магнитного момента электрона, упоминавшееся в предыдущей главе, с. 151). Однако она является весьма неупорядоченной (и не вполне непротиворечивой), так как изначально дает не имеющие физического смысла «бесконечные» ответы. Такие бесконечные значения, или расходимости, подлежат устранению с помощью так называемой процедуры «перенормировки». Не все квантовые теории поля поддаются перенормировке, и даже те, которые допускают перенормировку, наталкиваются на значительные вычислительные трудности.

Весьма популярен подход к квантовой теории поля через использование «интегралов по траекториям», включающих в себя образование квантовых линейных суперпозиций не только состояний различных частиц (как с помощью обычных волновых функций), но учитывающих все пространственно-временные истории физического поведения (доступный обзор см. в книге Фейнмана [1985]). Однако этот подход сам по себе приводит к дополнительным расходимостям, и придать смысл методу «интегралов по траекториям» можно только с помощью различных «математических трюков». Несмотря на несомненную силу и впечатляющую точность квантовой теории поля (в тех немногих случаях, когда теория может быть полностью применена), у физиков остается впечатление, что необходимо более глубокое понимание, прежде чем

---

[14] Однако между отдельным фотоном и электромагнитным полем существует важное различие в типе допустимых *решений* уравнения. Классические максвелловские поля с необходимостью *действительнозначные*, тогда как состояния фотона *комплекснозначные*. К тому же фотон должен удовлетворять так называемому условию «положительной частоты».

можно будет с уверенностью принять «картину физической реальности», к которой может привести квантовая теория поля [16].

Я хотел бы подчеркнуть, что согласие между квантовой теорией и специальной теорией относительности, достигающееся в квантовой теории поля, является лишь *частичным* — касается только U-части — и носит весьма формальный математический характер. Трудности непротиворечивой релятивистской интерпретации «квантовых скачков», связанных с R-частью, к которым приводят эксперименты типа ЭПР, даже не затрагиваются квантовой теорией поля. Кроме того, пока еще не существует непротиворечивой квантовой теории гравитационного поля, которой можно было бы верить. В главе 8 я выскажу некоторые догадки относительно того, что эти проблемы не могут быть никак не связанными между собой.

Кошка Шредингера

Наконец, обратимся к вопросу, который преследует нас с самого начала нашего описания. Почему мы не наблюдаем квантовых линейных суперпозиций объектов классических масштабов, например, крикетных шаров, находящихся одновременно в двух местах? Что заставляет определенные конфигурации атомов срабатывать как «измерительное устройство», так что R-процедура сменяет U? Разумеется, любая часть измерительного прибора сама по себе является частью физического мира и состоит из тех самых квантовомеханических компонент, поведение которых должен исследовать прибор. Почему бы не рассматривать измерительный прибор *вместе* с физической системой как *единую составную квантовую систему*? При таком подходе нет загадочного «внешнего» измерения. Составная система должна просто эволюционировать в соответствии с U. Но эволюционирует ли она именно так? Действие U-процедуры на составную систему полностью детерминистично и не оставляет места для вероятностных неопределенностей R-типа, встречающихся в «измерении» или «наблюдении», которые составная система производит над собой! В сказанном есть явное противоречие, которое проявляется особенно наглядно в знаменитом мысленном эксперименте, предложенном Эрвином Шредингером [1935]: в парадоксе «кошка Шредингера».

Представьте себе герметичный контейнер, спроектированный и построенный столь тщательно, что сквозь его стенки ни внутрь, ни наружу не проходит никакое физическое воздействие. Предположим, что внутри контейнера находится кошка, а также устройство, приводимое в действие («запускаемое») некоторым квантовым событием. Если это событие происходит, то устройство разбивает ампулу с синильной кислотой, и кошка погибает. Если событие не происходит, то кошка продолжает жить. В первоначальной версии Шредингера квантовым событием, запускающим устройство, был распад радиоактивного атома. Позвольте мне слегка модифицировать первоначальную версию Шредингера и выбрать в качестве квантового события, запускающего устройство, фотон, который, попадая в фотоэлемент, приводит его в действие — фотон, испущенный некоторым источником света в предопределенном состоянии и от-

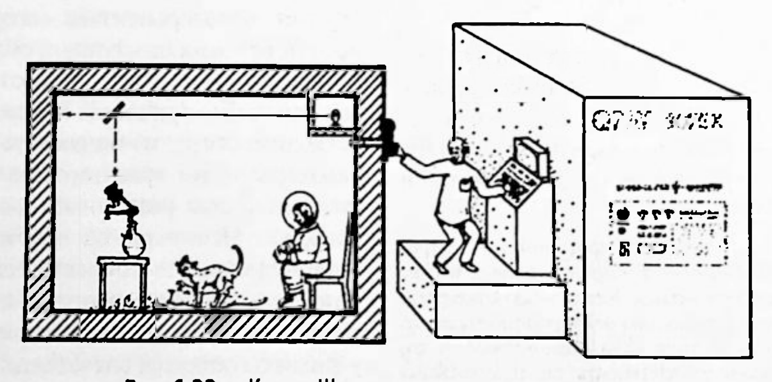


Рис. 6.33. «Кошка Шредингера» — с дополнениями

раженный от полупосеребренного зеркала (рис. 6.33). Отражение от зеркала расщепляет волновую функцию фотона на две отдельные части, одна из которых отражается, а другая проходит сквозь зеркало. Отраженная часть волновой функции фотона фокусируется на фотоэлементе так, что если фотон регистрируется фотоэлементом, то это означает, что он *отразился*. В этом случае синильная кислота выделяется и кошка погибает. Если же фотоэлемент *не срабатывает*, то это означает, что фотон *прошел* сквозь полупосеребренное зеркало до стенки контейнера, расположенной за зеркалом, и кошка осталась жива.

С точки зрения (довольно рискованного) наблюдателя, находящегося *внутри* контейнера, именно таким было бы описание событий, происходящих внутри контейнера. *Либо* считается, что фотон отразился, так как по свидетельству наблюдателя фотоэлемент зарегистрировал фотон, и кошка погибла, *либо* считается, что фотон прошел сквозь зеркало, так как по свидетельству наблюдателя фотоэлемент не зарегистрировал фотон, и кошка осталась жива. Либо одно, либо другое *действительно* происходит: реализуется **R**-процедура, и вероятность каждой возможности составляет 50% (потому что зеркало *полупосеребренное*). Но взглянем теперь на события с точки зрения наблюдателя, находящегося *снаружи* контейнера. Мы можем считать, что *начальный* вектор состояния содержимого контейнера был «известен» наблюдателю до того, как контейнер был герметически запечатан. (Я отнюдь не хочу сказать, что вектор состояния содержимого контейнера мог быть известен на практике, но ничто в квантовой теории не утверждает, что он не мог бы *в принципе* быть известен наблюдателю.) Согласно внешнему наблюдателю никакое «измерение» в действительности не производилось, поэтому вся эволюция вектора состояния должна была бы происходить в соответствии с **U**-процедурой. Фотон испускается источником в определенном состоянии (в этом оба наблюдателя сходятся во мнении), и его волновая функция расщепляется на две части с амплитудой $1/\sqrt{2}$ для каждой из частей (тогда квадрат модуля действительно даст вероятность 1/2). Так как все содержимое контейнера рассматривается внешним наблюдателем как одна квантовая система, линейная суперпозиция альтернатив должна выполняться вплоть до масштабов кошки. Существует амплитуда $1/\sqrt{2}$ того, что фотоэлемент зарегистрирует фотон, и амплитуда $1/\sqrt{2}$ того, что он фотон не зарегистрирует. *Обе* альтернативы должны быть представлены в состоянии и участвовать в квантовой линейной суперпозиции с равными весами. С точки зрения внешнего наблюдателя кошка есть не что иное, как линейная суперпозиция дохлой и живой кошек!

Убеждены ли мы в том, что в действительности все обстоит именно так? Сам Шредингер ясно и определенно заявил о том, что так не считает. Действительно, свое мнение он аргументировал тем, что **U**-процедура квантовой механики не должна применяться к чему-нибудь столь большому или столь сложному, как кошка. При попытке применить **U**-процедуру к столь большому и сложному объекту уравнение Шредингера где-то должно утратить силу. Разумеется, Шредингер имел право рассуждать так о своем собственном уравнении, но все остальные из нас лишены такой прерогативы! Наоборот, многие физики (в действительности большинство физиков) склонны считать, что в настоящее время имеется весьма много экспериментальных фактов, свидетельствующих в пользу **U**-процедуры, и нет ни одного экспериментального факта, который свидетельствовал бы против **U**, поэтому мы не имеем никакого права отказываться от этого типа эволюции даже на уровне кошки. Если принять эту точку зрения, то мы, кажется, будем вынуждены прийти к весьма *субъективному* представлению о физической реальности. Для внешнего наблюдателя кошка действительно есть не что иное, как линейная комбинация дохлой и живой кошек, и только когда контейнер, наконец, будет вскрыт, вектор состояния кошки коллапсирует в вектор одного из этих двух состояний. С другой стороны, для внутреннего наблюдателя (надлежащим образом защищенного от воздействия синильной кислоты) вектор состояния кошки коллапсировал бы гораздо раньше, и линейная

комбинация внешнего наблюдателя

$$|\psi\rangle = \frac{1}{\sqrt{2}}\{|\text{живая}\rangle + |\text{дохлая}\rangle\}$$

не имела бы смысла. Создается впечатление, что вектор состояния в конечном счете существует «только в воображении» наблюдателя!

Но можем ли мы принять такую субъективную точку зрения на вектор состояния? Предположим, что внешний наблюдатель не просто «заглядывает» в контейнер, а производит некую более изощренную процедуру. Предположим также, что, исходя из того, что он знает о начальном состоянии внутри контейнера, внешний наблюдатель сначала использует некоторый быстродействующий компьютер, чтобы на основании уравнения Шредингера *вычислить*, какое состояние действительно должно установиться внутри контейнера, и получить («правильный») ответ $|\psi\rangle$ (где $|\psi\rangle$ действительно включает в себя линейную суперпозицию дохлой кошки и живой кошки). Предположим далее, что внешний наблюдатель выполняет над содержимым контейнера *тот самый* эксперимент, который позволяет отличить состояние $|\psi\rangle$ от любого ортогонального ему состояния. (Как было показано выше, по правилам квантовой механики внешний наблюдатель *в принципе* может выполнить такой эксперимент, хотя осуществить его на практике было бы чрезвычайно трудно.) Вероятности двух исходов: «да, находится в состоянии $|\psi\rangle$» и «нет, находится в состоянии, ортогональном $|\psi\rangle$» — составляли бы, соответственно, 100% и 0%. В частности, для состояния $|\chi\rangle = |\text{дохлая}\rangle - |\text{живая}\rangle$, *ортогонального* $|\psi\rangle$, вероятность была бы равна 0. Невозможность состояния $|\chi\rangle$ в результате эксперимента может возникнуть только потому, что *обе* альтернативы $|\text{дохлая}\rangle$ и $|\text{живая}\rangle$ *сосуществуют* друг с другом.

То же самое можно было бы утверждать и в том случае, если бы мы подобрали соответствующим образом длины путей фотона (или плотность посеребренного слоя на поверхности зеркала), так чтобы вместо линейной суперпозиции состояний $|\text{дохлая}\rangle + |\text{живая}\rangle$ мы имели бы некоторую другую комбинацию, например, $|\text{дохлая}\rangle - i|\text{живая}\rangle$ и т. д. Все эти различные комбинации приводят к различным экспериментальным следствиям (в принципе!). Таким образом уже говорится не «просто» о некоторой форме сосуществования между жизнью и смертью, от которой зависит судьба нашей несчастной кошки. Допустимы все возможные *комплексные* комбинации, и все они (в принципе) отличимы одна от другой! Однако наблюдателю, находящемуся внутри контейнера, все эти комбинации представляются несущественными. Кошка *либо* жива, *либо* мертва. Каким образом мы можем придать смысл такого рода несоответствию? Я кратко приведу несколько различных точек зрения, высказанных по этому (и аналогичным) вопросу, хотя не подлежит сомнению, что я не смогу всем им дать равнозначную оценку.

Различные точки зрения на существующую квантовую теорию

Прежде всего практическая реализация эксперимента, аналогичного тому, который позволяет отличить состояние $|\psi\rangle$ от любого состояния, ортогонального $|\psi\rangle$, наталкивается на очевидные трудности. Не подлежит сомнению, что такой эксперимент *на практике* невозможен для внешнего наблюдателя. В частности, для этого внешнему наблюдателю понадобилось бы точно знать вектор состояния всего содержимого контейнера (включая наблюдателя, находящегося внутри контейнера), прежде чем он мог бы приступить к вычислению $|\psi\rangle$ в более поздние моменты времени! Однако мы требуем, чтобы такой эксперимент был невозможен не только на практике, но и *в принципе*, так как в противном случае у нас не было бы права изъять из физической реальности одно из состояний $|\text{живая}\rangle$ или $|\text{дохлая}\rangle$. Трудность заключается в том, что квантовая теория в том виде, в каком она существует сейчас, не дает никаких указаний относительно того, как должна быть проведена четкая линия между «возможными» и «невозможными» измерениями. Вполне вероятно, что такое четкое разграничение между теми и другими измерениями *должно было*

бы существовать. Но современная квантовая теория не позволяет провести такое разграничение. Чтобы провести разграничительную линию между «возможными» и «невозможными» измерениями, потребовалось бы *изменить* квантовую теорию.

Во-первых, нередко высказывают точку зрения, согласно которой все трудности исчезли бы, если бы мы адекватно учли «окружающую среду» интересующей нас системы. Действительно, *полностью* изолировать содержимое контейнера от внешнего мира практически невозможно. Как только окружающая среда начинает влиять на состояние содержимого контейнера, внешний наблюдатель не может считать, что состояние содержимого контейнера задается просто одним вектором состояния. Даже *собственное* состояние внешнего наблюдателя оказывается сложным образом коррелированным с состоянием содержимого контейнера. Кроме того, с внутренностью контейнера неразрывно связано огромное число различных частиц, эффекты различных возможных линейных комбинаций распространяются все дальше и дальше во вселенную, охватывая огромное число степеней свободы. Не существует *практического* способа (например, по наблюдению соответствующих эффектов интерференции), который позволил бы отличить эти комплексные линейные суперпозиции от вероятностно-взвешенных альтернатив. Это не должно быть связано просто с вопросом об изоляции содержимого контейнера от внешней среды. Сама кошка состоит из огромного числа частиц. Таким образом, комплексную линейную комбинацию дохлой кошки и живой кошки можно трактовать *как если бы* она была просто смесью вероятностей. Но лично я отнюдь не считаю такую трактовку удовлетворительной. Как и в предыдущем рассуждении, мы можем спросить, на какой стадии получение интерференционных эффектов официально объявляется «невозможным», в результате чего квадраты модулей амплитуд в комплексной суперпозиции могут быть объявлены вероятностными весами «дохлой» и «живой» кошки. Даже если «реальность» мира «в действительности» становится (в некотором смысле) *действительнозначным* вероятностным весом, каким образом это превращается в единственную альтернативу, ту или иную? Я не усматриваю, каким образом *реальность* может трансформироваться из комплексной (или действительной) линейной *суперпозиции* двух альтернатив в *одну или другую* из этих альтернатив на основе одной лишь эволюции U. Мне кажется, что подобный взгляд возвращает нас к субъективной точке зрения на мир.

Иногда высказывают мнение, что сложные системы должны в действительности описываться не «состояниями», а их обобщением, получившим название *матриц плотности* (фон Нейман [1955]). Последние включают в себя и классические вероятности и квантовые амплитуды. В этом случае для описания реальности берутся много квантовых состояний. Матрицы плотности полезны, но сами по себе они не решают глубоко проблематичные вопросы квантового измерения.

Можно попытаться придерживаться той точки зрения, что реальная эволюция — это детерминистский U-процесс, а вероятности возникают из-за неопределенностей в нашем знании того, что *в действительности* представляет собой квантовое состояние сложной системы. Такая точка зрения очень близка к «классическому» взгляду на происхождение вероятностей, согласно которому вероятности возникают из неопределенностей в начальном состоянии. Можно представить, что крохотные различия в начальном состоянии могут привести к огромным различиям в эволюции, таким, как «хаос», который встречается у классических систем (см., например, о предсказании погоды в главе 5, с. 166). Однако такие «хаотические» эффекты просто не могут возникнуть в рамках действия одной лишь U-процедуры, так как U *линейна*: нежелаемые линейные суперпозиции остаются таковыми навсегда при действии U! Чтобы выделить из нежелаемой суперпозиции ту или иную альтернативу, требуется нечто *нелинейное*, поэтому самой U-процедуры для этого недостаточно.

Можно придти к другой точке зрения, если заметить, что единственное очевидное несоответствие с наблюдением в эксперименте с кошкой Шредингера возникает,

по-видимому, потому, что имеются *сознательные наблюдатели* один (или два!) внутри и один снаружи контейнера. Возможно, законы комплексной квантовой линейной суперпозиции *неприменимы* к сознанию! Грубая математическая модель, отражающая эту точку зрения, была предложена Эугеном П. Вигнером [1961]. Вигнер предположил, что линейность уравнения Шредингера может нарушаться для существ, наделенных сознанием (или просто «живых» существ), и это уравнение подлежит замене на некоторую нелинейную процедуру, согласно которой та или иная из альтернатив должна быть отброшена. Читателю может показаться, что поскольку я пытаюсь выяснить, какого рода роль могут играть в нашем сознании квантовые явления (а я действительно пытаюсь это выяснить), я должен был бы принять такую точку зрения как очень желательную. Однако она меня совсем не удовлетворяет. Мне кажется, что она приводит к весьма одностороннему и искаженному взгляду на *реальность* окружающего мира. Те уголки вселенной, где обитает сознание, могут быть весьма малочисленными и разделенными огромными расстояниями. С рассматриваемой точки зрения, *только* в этих редких и разбросанных на большие расстояния уголках комплексные квантовые линейные суперпозиции могут быть расщеплены на реальные альтернативы. Возможно, что для *нас* такие особые уголки выглядели бы так же, как и остальная вселенная, так как куда бы мы ни посмотрели (или что бы ни наблюдали каким-либо другим способом), объект наблюдения в силу самого акта нашего сознательного наблюдения оказался бы «расщепленным на альтернативы», независимо от того, *было ли или не было* такое расщепление произведено до нашего наблюдения. Столь сильная односторонность могла бы привести к весьма искаженной картине *реальности* нашего мира, и я со своей стороны принял бы ее весьма неохотно.

Существует другая точка зрения, связанная в чем-то с предыдущей, которая сводит роль сознания к другому (противоположному) пределу. Она была выдвинута Джоном Уилером [1938] и получила название *соучаствующей* (*партисипаторной*) *вселенной*. Отметим, например, что эволюция сознательной жизни на нашей планете обусловлена подходящими мутациями, происходившими в различное время. Предположительно это были квантовые события, поэтому они могли бы существовать только в виде линейной суперпозиции до тех пор, пока они не довели эволюцию до мыслящих существ, самое существование которых зависит от всех «правильных» мутаций, имевших место в действительности! Именно наше присутствие, согласно этой идее, вызывает к существованию наше прошлое. Парадоксальность, присущая этой картине, может вызвать определенный интерес, но я лично вижу в ней много проблем и не считаю правдоподобной.

Другая точка зрения, также по-своему логичная, но приводящая к не менее странной картине — так называемая теория *множественности миров*, впервые выдвинутая Хью Эвереттом III [1957]. Согласно этой теории R-процедура вообще не имеет места. Вся эволюция вектора состояния (который считается реалистическим) все время управляется детерминистской U-процедурой. Отсюда следует, что несчастная кошка Шредингера вместе с облаченным в защитный костюм наблюдателем внутри контейнера действительно должны существовать в некоторой комплексной линейной комбинации, причем кошка должна представлять собой некоторую суперпозицию живой и дохлой. Однако дохлое состояние коррелировано с одним состоянием сознания наблюдателя, находящегося внутри контейнера, а живое состояние — коррелировано с другим состоянием его сознания (и, частично, с сознанием кошки, а в конечном счете и с состоянием сознания внешнего наблюдателя, после того, как содержимое контейнера открывается его наблюдению). Состояние каждого наблюдателя, с точки зрения Эверетта, надлежит считать «расщепляющимся», так как наблюдатель теперь как бы существует в двух экземплярах, причем каждый из экземпляров обладает различным жизненным опытом (один видит кошку живой, другой — дохлой). В действительности не только наблюдатель, но и весь мир, в котором он обитает, расщепляется на два мира (или на большее число миров) при каждом измерении,

производимом им над окружающим миром. Такое расщепление повторяется снова и снова — не только из-за измерений, производимых наблюдателями, но и из-за усиления до макроскопических масштабов квантовых событий, вследствие чего «ветви» этого мира чудовищно множатся. Действительно, каждая альтернативная возможность сосуществовала бы в некоторой огромной суперпозиции. Вряд ли теория множественности миров — самая экономичная точка зрения, но мои собственные возражения связаны отнюдь не с отсутствием экономичности. В частности, я не понимаю, почему сознание непременно должно быть осведомлено только об «одной» из альтернатив в некоторой линейной суперпозиции. Что такое в сознании настоятельно требует, чтобы мы не могли быть «осведомлены» о дразнящей линейной комбинации дохлой и живой кошек? Мне кажется, что необходимо разработать теорию сознания, прежде чем теорию множественности миров удастся обтесать, чтобы она согласовывалась с реальными наблюдениями. Я не вижу, какая взаимосвязь существует между «истинным» (объективным) вектором состояния вселенной и тем, что, как предполагается, мы «наблюдаем». Высказывались мнения, будто в такой картине можно эффективно вывести «иллюзию» **R**-процедуры, но я не думаю, что подобные утверждения соответствуют истине. В конце концов, для того, чтобы описанная выше схема заработала, необходимы еще некоторые дополнительные компоненты. Мне кажется, что теория множественности миров привносит сама по себе множество новых трудностей, не затрагивая по-настоящему реальные загадки квантового измерения (см. Де Витт, Грэхем [1973]).

К чему мы пришли после всего сказанного?

Затронутые выше вопросы в том или ином обличье присутствуют в *любой* интерпретации квантовой механики — в том виде, в каком эта теория существует в настоящее время. Приведем краткий обзор того, что стандартная квантовая теория в действительности говорит нам о том, каким образом мы должны описывать мир, особенно в отношении этих удивительных вопросов, и затем спросим: куда мы намерены двигаться дальше?

Прежде всего напомним, что описания, даваемые квантовой теорией, по-видимому, разумно (полезно?) применимы только на так называемом *квантовом уровне* — молекул, атомов или субатомных частиц, а также на бóльших масштабах при условии, что разности энергии между альтернативными возможностями остаются очень малыми. На квантовом уровне мы должны рассматривать такие «альтернативы» как нечто способное *сосуществовать* в виде суперпозиции с комплексными коэффициентами. Используемые в качестве весов комплексные числа называются *амплитудами вероятности*. Каждая из совокупности различных альтернатив с комплексными коэффициентами определяет свое, отличное от других, *квантовое состояние*, и любая квантовая система должны допускать описание таким квантовым состоянием. Нередко (наиболее ярко это проявилось в примере со *спином*) бывает и так, что нам нечего сказать относительно того, каковы должны быть «реальные» альтернативы, образующие квантовое состояние, и каковы должны быть всего лишь «комбинации» альтернатив. В любом случае пока система *остается* на квантовом уровне, квантовое состояние эволюционирует полностью *детерминистским* образом. Эта детерминистская эволюция и есть **U**-процесс, управляемый важным *уравнением Шредингера*.

Когда эффекты различных квантовых альтернатив оказываются увеличенными до *классического уровня*, так что различия между альтернативами становятся столь большими, что мы можем воспринимать их непосредственно, тогда такие суперпозиции с комплексными коэффициентами, по-видимому, перестают существовать. Вместо этого надо образовывать квадраты модулей комплексных амплитуд (т. е. брать квадраты их расстояний до начала координат на комплексной плоскости), и эти *действительные* числа теперь играют роль настоящих *вероятностей* для рассматриваемых альтернатив. В реальности физического эксперимента в соответствии с **R**-процедурой

(называемой редукцией вектора состояния, или коллапсом волновой функции; полностью отличной от U) выживает только *одна* из альтернатив. Именно здесь и только здесь в игру вступает индетерминизм квантовой теории.

Можно серьезно обосновать тезис о том, что квантовое состояние дает *объективную* картину. Но эта картина может быть сложной и даже парадоксальной. Когда в процессе участвуют несколько частиц, квантовые состояния могут становиться (и обычно *становятся*) очень сложными. Индивидуальные частицы не имеют своих собственных «состояний», а существуют только в сложных взаимосвязях с другими частицами, называемых *корреляциями.* Когда частица «наблюдается» в одной области в том смысле, что она «запускает» какой-то эффект, который затем увеличивается до классического уровня, после этого должна вступить в действие R-процедура, а это, по-видимому, оказывает *одновременно* влияние на все другие частицы, коррелированные с данной частицей. Эксперименты типа Эйнштейна — Подольского — Розена (ЭПР) (например, эксперимент Аспекта, в котором квантовый источник испускает в противоположных направлениях два фотона, а затем, когда фотоны оказываются на расстоянии нескольких метров друг от друга, производится порознь измерение их поляризаций) выявляют четкую наблюдательную суть озадачивающего, но существенного факта квантовой физики: она *нелокальна* (и поэтому фотоны в эксперименте Аспекта не могут рассматриваться как отдельные независимые сущности)! Если считать, что R-процедура действует объективно (а именно это, насколько можно судить, должно следовать из объективности квантового состояния), то тем самым нарушается дух специальной теории относительности. По-видимому, *не существует объективного пространственно-временнóго* описания (редуцируемого) вектора состояния, которое не противоречило бы требованиям специальной теории относительности! Однако *наблюдательные* эффекты квантовой теории не нарушают требований специальной теории относительности.

Квантовая теория умалчивает о том, *когда* и *почему* в действительности (или в воображении?) должна иметь место R-процедура. Кроме того, сама по себе R-процедура не дает надлежащего объяснения, почему мир на классическом уровне «выглядит» классическим. «Большинство» квантовых состояний совсем не похожи на классические состояния!

К чему мы пришли после всего сказанного? Я убежден, что необходимо вполне серьезно рассматривать возможность того, что квантовая механика просто *неверна*, когда ее применяют к макроскопическим телам, или, точнее, что законы U и R дают только превосходные приближения к некоторой более полной, но еще не разработанной теории. И лишь *комбинация* законов U и R, но не «законы U» в отдельности, дает все то чудесное согласие с наблюдением, которым так радует существующая ныне теория. Если бы линейность U-процедуры допускала распространение на макроскопический мир, то мы должны были бы принять как физическую реальность комплексные линейные комбинации различных пространственных положений (или различных спинов и т.д.) крикетных шаров и тому подобных макроскопических объектов. Но здравый смысл говорит нам, что мир в действительности ведет себя не так! Крикетные шары действительно могут быть хорошо аппроксимированы описаниями *классического* мира. Крикетные шары обладают разумно хорошо определенными положениями в пространстве, и их нельзя видеть в двух местах одновременно, как это разрешают линейные законы квантовой механики. Если U и R-процедуры подлежат замене каким-то более широким законом, то в отличие от уравнения Шредингера, этот новый закон должен быть *нелинейным* (так как R-процедура сама действует нелинейно). Некоторые люди возражают против такого утверждения, совершенно справедливо ссылаясь на то, что глубокое математическое изящество стандартной квантовой теории во многом обусловлено ее линейностью. Однако я считаю, что было бы удивительно, если бы квантовая теория в будущем не претерпела бы некоторых фундаментальных изменений, преобразуясь в такую

теорию, к которой линейный вариант стандартной квантовой механики был бы всего лишь приближением. Примеры тому уже имеются. Созданная Ньютоном изящная и мощная теория всемирного тяготения во многом опиралась на то обстоятельство, что силы тяготения суммируются *линейно.* Но с появлением общей теории относительности Эйнштейна стало ясно, что эта линейность всего лишь приближение (хотя и превосходное), и что изящество теории Эйнштейна превосходит даже изящество теории Ньютона!

Я никогда не скрывал свое убеждение, что решение загадок квантовой теории должно лежать в построении усовершенствованной теории. И хотя подобная точка зрения не общепринята, но также и не совсем отвергнута. (Ее придерживались многие из основателей квантовой теории. Я уже изложил взгляды Эйнштейна. Шредингер [1935], де Бройль [1956] и Дирак [1939] также рассматривали стандартную квантовую механику как неокончательную теорию.) Но даже если некто убежден в необходимости каким-то образом модифицировать квантовую теорию, ограничения на то, *каким образом* может быть произведена такая модификация, оказываются весьма жесткими. Возможно, что в конце концов приемлемой сочтут точку зрения, связанную с использованием «скрытых переменных». Но нелокальность, столь наглядно проявившаяся в экспериментах типа ЭПР, бросает суровый вызов любому «реалистическому» описанию мира, которое может комфортно вписаться в обычное пространство-время — пространство-время того особого типа, которое было дано нам в соответствии с принципами специальной теории относительности, — поэтому я убежден в необходимости гораздо более радикальных изменений. Кроме того, между квантовой механикой и экспериментом не было обнаружено никаких расхождений, если не рассматривать, разумеется, явное отсутствие линейной суперпозиции крикетных шаров как контраргумент. Мое личное мнение сводится к тому, что несуществование линейных суперпозиций крикетных шаров действительно *является* контраргументом! Но само по себе от этого не много толку. Мы знаем, что на субмикроскопическом уровне квантовые законы действительно работают; но на уровне крикетных шаров действует классическая физика. Где-то между ними находится закон, который нам необходимо понять, чтобы увидеть, каким образом квантовый мир возникает внутри классического мира. Кроме того, я убежден, что этот новый закон нам непременно понадобится, если мы собираемся понять, как функционирует наш разум! А для всего этого, по моему глубокому убеждению, нам необходимо искать новые подходы.

В своем изложении квантовой теории в этой главе я придерживался всецело традиционной точки зрения, хотя, может быть, сильнее, чем обычно, подчеркивал геометрический и «реалистический» аспекты. В следующей главе мы уделим особое внимание поиску недостающих ключевых моментов — того, что, по моему глубокому убеждению, должно нам дать указания на то, какой должна быть усовершенствованная квантовая механика. Наше путешествие начнется у порога нашего дома, но затем мы будем вынуждены значительно удалиться от него. Оказывается, что нам необходимо исследовать весьма далекие области пространства и обратиться даже к самому началу времен!

Примечания

1. Я считаю само собой разумеющимся, что любая «серьезная» философская точка зрения должна содержать по крайней мере изрядную долю реализма. У меня всегда вызывает удивление, когда я узнаю о серьезных мыслителях — нередко физиках, рассматривающих следствия, к которым приводит квантовая механика, — которые занимают сильно субъективную точку зрения, согласно которой в действительности *никакого* реального мира — «там, вовне» — вообще нет! То, что я придерживаюсь где только возможно реалистической линии, отнюдь не означает, что мне неизвестно о том, с какой

серьезностью отстаиваются подобные субъективные взгляды, — но я просто не могу придать им смысл. Тех, кто желает ознакомится с мощной и занимательной атакой на субъективизм такого рода, я приглашаю читать книгу Гарднера [1983], глава 1.

2. В частности, Дж. Дж. Бальмер отметил в 1895 году, что частоты спектральных линий водорода удовлетворяют формуле $R(n^{-2} - m^{-2})$, где n и m — положительные целые числа (R — постоянная).

3. Возможно, нам не следовало бы слишком легко отказываться от этой «чисто полевой» картины. Эйнштейн, который (как мы увидим в дальнейшем) глубоко сознавал дискретный характер квантовых частиц, провел последние тридцать лет своей жизни, пытаясь построить более общую теорию такого классического типа. Но попытки Эйнштейна, как и все прочие попытки, оказались тщетными. По-видимому, для объяснения дискретной природы частиц необходимо что-то еще помимо классического поля.

4. Эти две эволюционные процедуры были описаны в классическом труде выдающегося американского математика венгерского происхождения Джона (Яноша) фон Неймана [1955]. Его «процесс 1» — то, что я назвал R-процедурой — «редукцией вектора состояния», а его «процесс 2» — то, что я назвал U-процедурой — «унитарной эволюцией» (унитарность означает, что амплитуды вероятности в ходе эволюции сохраняются). На самом деле существуют и другие (хотя и эквивалентные) описания эволюции U квантового состояния, в которых не используется термин «уравнение Шредингера». Например, в *«картине Гейзенберга»* состояние описывается таким образом, что кажется, будто оно вообще не эволюционирует; динамическая эволюция понимается как непрерывный сдвиг системы координат положения/импульса. Разные отличия этих картин для нас сейчас несущественны, так как описания процесса U полностью эквивалентны.

5. Для полноты следовало бы также привести все требуемые алгебраические правила, записанные в используемых в тексте обозначениях (Дирака),

$$|\psi\rangle + |\chi\rangle = |\chi\rangle + |\psi\rangle,$$
$$|\psi\rangle + (|\chi\rangle + |\varphi\rangle) = (|\psi\rangle + |\chi\rangle) + |\varphi\rangle,$$
$$(z + w)|\psi\rangle = z|\psi\rangle + w|\psi\rangle,$$
$$z\,(|\psi\rangle + |\chi\rangle) = z|\psi\rangle + z|\chi\rangle,$$
$$z\,(w|\psi\rangle) = (zw)|\psi\rangle,$$
$$1|\psi\rangle = |\psi\rangle,$$
$$|\psi\rangle + 0 = |\psi\rangle,$$
$$0|\psi\rangle = 0 \quad \text{и} \quad z0 = 0.$$

6. Существует важная операция, называемая *скалярным* произведением (или внутренним произведением) двух векторов, которая может быть использована для того, чтобы очень просто выразить такие понятия, как «единичный вектор», «ортогональность» и «амплитуда вероятности». (В обычной векторной алгебре скалярное произведение равно $ab\cos\vartheta$, где a и b — длины векторов, а ϑ — угол между их направлениями.) Скалярное произведение векторов из гильбертова пространства дает *комплексное* число. Скалярное произведение двух векторов состояния $|\psi\rangle$ и $|\chi\rangle$ записывается в виде $\langle\psi|\chi\rangle$. Для него справедливы алгебраические правила $\langle\psi|(|\chi\rangle + |\varphi\rangle) = \langle\psi|\chi\rangle + \langle\psi|\varphi\rangle$, $\langle\psi|(q|\chi\rangle) = q\langle\psi|\chi\rangle$ и $\langle\psi|\chi\rangle = \overline{\langle\chi|\psi\rangle}$, где черта сверху означает комплексное сопряжение. (Числом, комплексно сопряженным с $z = x + iy$, называется $\bar{z} = x - iy$, где x и y — действительные числа; обратите внимание на то, что $|z|^2 = z\bar{z}$.) Ортогональность векторов состояния $|\psi\rangle$ и $|\chi\rangle$ записывается в виде соотношения $\langle\psi|\chi\rangle = 0$. Квадрат длины вектора состояния $|\psi\rangle$ есть величина $|\psi|^2 = \langle\psi|\bar{\psi}\rangle$, поэтому условие нормировки $|\psi\rangle$ к единичному вектору представимо в виде $\langle\psi|\bar{\psi}\rangle = 1$. Если «акт измерения» вызывает скачкообразный переход состояния $|\psi\rangle$ либо в состояние $|\chi\rangle$, либо во что-то, ортогональное $|\chi\rangle$, то амплитуда этого скачкообразного перехода в состояние $|\chi\rangle$ равна $\langle\chi|\psi\rangle$ в предположении, что $|\psi\rangle$ и $|\chi\rangle$ нормированы. Без нормировки вероятность скачкообразного перехода из $|\psi\rangle$ в $|\chi\rangle$ можно представить в виде $\langle\chi|\psi\rangle\langle\psi|\chi\rangle/\langle\chi|\chi\rangle\langle\psi|\psi\rangle$. (См. Дирак [1947].)

7. Для тех, кто знаком с операторным формализмом квантовой механики, это измерение (в обозначениях Дирака) определяется ограниченным эрмитовым оператором $|\chi\rangle\langle\chi|$. Собственное значение 1 (для нормированного $|\chi\rangle$) означает ДА, а собственное значение 0 — НЕТ. (Векторы $\langle\chi|$, $\langle\psi|$ и т. д. принадлежат гильбертову пространству, *дуальному* к исходному.) См. фон Нейман [1955], Дирак [1947].

8. В предыдущем описании квантовой системы, состоящей из одной частицы, я прибег к сверхупрощению, проигнорировав спин и предположив, что состояние может быть описано заданием одного лишь простран-

ственного положения. Действительно, существуют некоторые частицы, называемые *скалярными*, их примерами могут служить ядерные частицы, известные под названием *пионов* (π-мезоны, см. с. 201), или некоторые атомы, для которых спин оказывается равным нулю. Для таких (и только для таких) частиц приведенное выше описание в терминах одного лишь пространственного положения действительно будет достаточным.

9. Пусть $|\swarrow\rangle = \bar{z}|\uparrow\rangle - \bar{w}|\downarrow\rangle$, где $\bar{z}$ и $\bar{w}$ — комплексно сопряженные чисел z и w. (См. прим. 6.)

10. Существует стандартная экспериментальная установка, известная как прибор Штерна—Герлаха, которую можно использовать для измерения спинов атомов. Атомы выпускаются в пучок, который проходит в сильно неоднородном магнитном поле, направление неоднородности которого задает направление, в котором производится измерение спина. Пучок расщепляется на два (для атома со спином 1/2 или на большее число частей — для атома с бо́льшим спином), один пучок дает атомы с ответом ДА на измерение спина, а другой — атомы с ответом **НЕТ** на измерение спина. К сожалению, по некоторым техническим причинам, не имеющим отношения к интересующим нас вопросам, такой прибор не может быть использован для измерения спина электрона, и поэтому приходится прибегать к косвенной процедуре (см. Мотт, Мэсси [1965]). По этой и по другим причинам я предпочитаю не вдаваться в подробности относительно того, как в действительности измеряют спин электрона.

11. Пытливый читатель может самостоятельно проверить геометрию, приведенную в тексте. Проще всего, если мы сориентируем сферу Римана так, чтобы α-направление было направлением «вверх», а β-направление лежало в плоскости, натянутой на направления «вверх» и «вправо», т. е. задаваемой параметром $q = \operatorname{tg}(\vartheta/2)$ на сфере Римана, а затем воспользуемся формулой $\langle\chi|\psi\rangle\langle\psi|\chi\rangle/\langle\chi|\chi\rangle\langle\psi|\psi\rangle$ для вероятности перехода скачком из $|\psi\rangle$ в $|\chi\rangle$. (См. прим. 6.)

12. Математически можно сказать, что пространство двухчастичных состояний есть *тензорное произведение* пространства состояний первой частицы и пространства состояний второй частицы. Таким образом, $|\chi\rangle|\psi\rangle$ есть тензорное произведение состояний $|\chi\rangle$ и $|\psi\rangle$.

13. Блестящий австрийский физик Вольфганг Паули, сыгравший выдающуюся роль в развитии квантовой механики, выдвинул свой принцип запрета в 1925 году в качестве гипотезы. Полная квантовомеханическая теория того, что мы ныне называем «фермионами», была разработана в 1926 году выдающимся физиком Энрико Ферми и великим Полем Дираком, с которым мы уже несколько раз встречались по ходу изложения. Статистическое поведение фермионов соответствует «статистике Ферми — Дирака» (отличной от «статистики Больцмана» — классической статистики различимых частиц). «Статистика Бозе — Эйнштейна» бозонов была разработана для рассмотрения фотонов замечательным индийским физиком Шатьендранатом Бозе и Альбертом Эйнштейном в 1924 году.

14. Это настолько замечательный и важный результат, что стоит изложить еще один его вариант. Предположим, что существуют всего лишь *две* настройки для E-измерителя: вверх [↑] и вправо [→], и *две* настройки для P-измерителя — под углом 45° к направлению вправо вверх [↗] и под углом 45° к направлению вправо вниз [↘]. Предположим, что *реальные* настройки для E- и P-измерителей — соответственно [→] и [↗]. Тогда вероятность того, что E- и P-измерения дадут согласующиеся результаты, равна $(1/2)(1 + \cos 135°) = 0{,}146\ldots$, что чуть меньше 15%. Длинная последовательность экспериментов при таких настройках, например,

E: ДННДНДДДНДДННДННННДДН...
P: НДДНННДНДННДДНДДНДННД...
✓ ✓ ✓

даст нам согласие лишь немного меньше 15%. Предположим теперь, что на P-измерения никак не влияет E-настройка — т. е. что *если* E-настройка была бы [↑], а не [→], то исходы P-измерений были бы такими же, а так как угол между [↑] и [↗] такой же, как между [→] и [↗], то вероятность согласия между исходами P-измерений и новых E-измерений (обозначим их, например, E'-измерениями) по-прежнему была бы лишь немного меньше 15%. С другой стороны, если E-настройка была бы [→], как прежде, а P-настройка была бы [↘], а не [↗], то серия E-результатов осталась бы такой же, как прежде, а новая серия P-результатов, которую мы обозначим, например, P', была бы в согласии лишь немногим меньше 15% с исходной серией E-результатов. Отсюда следует, что согласие между P'-измерением и E'-измерением могло бы

быть не выше 45% (=15%+15%+15%), если бы эти измерения производились бы, соответственно, при настройках [↘] и [↑]. Но угол между [↘] и [↑] равен 135°, а не 45°, поэтому вероятность согласия *должна* была бы быть чуть больше 85%, а не 45%. Это — противоречие, показывающее, что допущение, согласно которому выбор измерения, произведенного E-измерителем, не может влиять на результаты P-измерений (и *наоборот*) должно быть ложно! За этот пример я признателен Дэвиду Мермину. Вариант, приведенный в тексте, заимствован из его статьи (Мермин [1985]).

15. Более ранние результаты, принадлежавшие Фридману и Клаузеру [1972], основаны на идеях, высказанных Клаузером, Хорном, Шимони и Холтом [1969]. В этих экспериментах все еще имеется один спорный пункт в связи с тем, что используемые в экспериментах детекторы фотонов обладают КПД, существенно меньшим 100%, поэтому лишь сравнительно малая доля испущенных фотонов оказывается реально детектированной. Однако даже с такими детекторами согласие с квантовой теорией столь совершенно, что трудно понять, как повышение КПД детекторов способно внезапно *ухудшить* согласие с теорией!

16. Кажется, что квантовая теория поля дает некоторый простор для невычислимости. (См. Комар [1964].)

Глава 7

Космология и стрела времени

Течение времени

Главным для нашего осознания бытия является ощущение движения времени. Нам *кажется*, что мы всегда движемся вперед, из определенного прошлого в неопределенное будущее. Мы чувствуем, что прошлое позади, и с ним уже ничего нельзя поделать. Оно неизменно и, в определенном смысле, уже «не с нами». То, что мы знаем в данный момент о прошлом, может брать начало в наших записях, рождаться из наших воспоминаний или подтверждаться иными свидетельствами — но в любом случае мы не склонны подвергать сомнению *реальность* прошлого. Прошлое является для нас чем-то совершенно определенным и не может стать (сейчас) другим. Что было — то было, и теперь уже ни мы, ни кто-либо другой не в силах ничего изменить! Будущее, с другой стороны, выглядит еще неопределенным. Оно может проявить себя и так и этак. Возможно, что этот «выбор» полностью определен физическими законами; а, возможно, отчасти и нашими собственными решениями (или Богом) — но в любом случае этот выбор, *кажется*, еще только предстоит сделать. То, что есть в данный момент — это всего-лишь *потенциальные возможности*, один из вероятных вариантов будущего.

Каждый миг, когда мы осознаем течение времени, наиболее близкая часть необозримого и кажущегося неопределенным будущего непрерывно превращается в настоящее, и, таким образом, добавляет свою строку в анналы прошлого. Иногда нам кажется, что *мы сами* были лично «ответственны» за какое-то действие, повлиявшее на конкретный выбор именно того возможного будущего, которое реализовалось на самом деле и стало необратимым в реальности прошлого. Но чаще, мы чувствуем себя пассивными наблюдателями — впрочем, весьма благодарными за такое освобождение от ответственности — того, как неумолимо свершившееся прошлое расширяет свои рамки за счет неопределенного будущего.

Как ни странно, физика рисует нам сегодня совершенно другую картину. Все основные уравнения физики симметричны во времени. Они оказываются одинаково справедливыми как для одного направления времени, так и для другого. Будущее и прошлое, с точки зрения физики, совершенно равноправны. Законы Ньютона, уравнения Гамильтона, уравнения Максвелла, общая теория относительности Эйнштейна, уравнение Дирака, уравнение Шредингера — все они, в действительности, остаются неизменными при обращении направления времени (т. е. замены координаты t, представляющей время, на $-t$). Вся классическая механика, вместе с **U**-частью квантовой, полностью обратима во времени. Может возникнуть вопрос о том, обратима ли **R**-часть квантовой механики. Подробно мы обсудим это в следующей главе, а пока давайте ограничимся соображениями «здравого смысла» и будем считать, что операция **R**, несмотря на свой вид, также должна быть взята симметричной во времени (см. Ааронов,

Бергманн, Лебовиц [1964]). В этом случае нам, по всей видимости, придется продолжить где-нибудь в другом месте поиски той области, в которой наши физические законы устанавливали бы различие между прошлым и будущим.

Перед тем, как отправиться в путь, имеет смысл немного задержаться еще на одном озадачивающем несоответствии между нашим субъективным восприятием времени и представлениями современной физики. Дело в том, что согласно специальной теории относительности, такого понятия, как «сейчас», на самом деле вообще не существует. Из того, что мы имеем в этой теории, наилучшим приближением к нему было бы «пространство одновременных событий» наблюдателя в пространстве-времени, показанном на рис. 5.21 на с. 187 — но оно, однако, зависит от *движения* наблюдателя! «Сейчас» для одного наблюдателя совсем не то же самое, что «сейчас» для другого [1]. Исследуя два события A и B в пространстве времени, один наблюдатель (U) может заключить, что событие B лежит в фиксированном прошлом, а событие A — в неопределенном будущем; в то время как для второго наблюдателя (V), A может оказаться в фиксированном прошлом, а B — в неопределенном будущем (рис. 7.1)! Мы не можем утверждать, что какое-либо из событий A или B остается неопределенным, в то время как другое из них уже определено.

Вернемся к рассуждениям на с. 188 и рис. 5.22. Два человека разминулись на улице; для одного из них космическая флотилия



Рис. 7.1. Может ли время действительно «течь»? Для наблюдателя U, B может находиться в «фиксированном» прошлом, в то время как A лежит еще в «неопределенном» будущем. Наблюдатель V придерживается противоположной точки зрения!

Андромеды уже отправилась в путешествие, в то время как для другого решение о том, состоится путешествие или нет, еще даже не принято. Возможно ли это? Ведь если хотя бы *один* из людей уже знает, что решение было принято, тогда, казалось бы, никакой неопределенности здесь быть *не может*. Запуск космической флотилии — реальность. На самом деле, ни один из этих людей в момент наблюдения еще не может что-либо *знать* о запуске. Они узнают о нем позднее, когда наблюдения с Земли подтвердят, что флотилия действительно уже в пути. Тогда они могут еще раз сопоставить свои прошлые наблюдения [2] и прийти к заключению, что *во время наблюдения* для одного из них решение о запуске лежало в неопределенном будущем, тогда как для другого, — в определенном прошлом. Имеет ли смысл, *в таком случае*, говорить о какой-либо неопределенности будущего? А может быть будущее для них *обоих* было уже изначально «фиксированным»?

Складывается впечатление, будто всякая определенность чего бы то ни было неизбежно приводит к определенности пространства-времени в целом! В этом случае вовсе нет никакого «неопределенного» будущего. *Все* пространство-время должно быть изначально фиксированным и никакой неопределенности просто нет места. Кажется, именно так думал и сам Эйнштейн (см. Пайс [1982], с. 444). Следуя этой логике, можно заключить, что нет и течения времени. Остается только «пространство-время», в котором нет места будущему, в чьи «владения» неумолимо вторгается определенное прошлое! (Читатель может в этом месте задаться вопросом о роли квантово-механических «неопределенностей». Я вернусь к вопросам, навеянным квантовой механикой, в следующей главе. Сейчас будет лучше проводить все рассуждения в рамках чисто классической картины.)

Мы видим, что налицо впечатляющие несоответствия между нашим субъективным ощущением потока времени и тем, как представляют нам физическую реальность наши (удивительно точные) теории. Эти несоответствия, скорее всего, свидетельствуют о существовании иных принципов, которые, по-видимому, и должны лежать глубоко

в основе наших субъективных ощущений — предполагая (как мне кажется), что эти принципы могут быть адекватно выражены на языке некоторой физической теории. Во всяком случае, представляется бесспорным, что какая бы теория ни работала, она должна нести в себе существенно асимметричную во времени составляющую, т. е. должна, так или иначе, отделять прошлое от будущего.

Но если уравнения физики никак не различают, как кажется, прошлое и будущее, и если даже сама идея «настоящего» так плохо согласуется с относительностью — тогда в какой же части мироздания нам следует искать ту область, где физические законы в большей степени соответствуют нашему восприятию мира? К счастью, если признаться честно, несоответствия не столь уж катастрофичны, как могло бы показаться. Наша физическая картина мира и в самом деле содержит некоторые фундаментальные составляющие, отличные от простых эволюционных уравнений, и при этом некоторые из них действительно несут в себе временну́ю асимметрию. Наиболее важная из этих составляющих носит название «второго начала термодинамики». Давайте попробуем разобраться, о чем в данном случае идет речь.

Неумолимое возрастание энтропии

Представим себе стакан воды, стоящий на самом краю стола. Если его слегка подтолкнуть, он, скорее всего, упадет на пол, наверняка разобьется вдребезги на множество осколков, а вода расплескается повсюду, возможно, частично поглотившись ковром или просочившись в щели между половицами. Наш стакан воды в этой ситуации лишь добросовестно следует уравнениям физики. Ньютоновское описание оказывается справедливым здесь в полной мере. Каждый из атомов в стекле и в воде подчиняется законам Ньютона (рис. 7.2). А теперь попробуем прокрутить эту картину в обратном направлении. В силу обратимости во времени законов Ньютона, вода могла бы также легко истечь из ковра и из щелей в половицах, заполнить стакан, который в это время ловко собирал бы себя из множества отколовшихся осколков, а затем все это могло запрыгнуть на высоту стола и устроиться в равновесии на его краю. И все это, так же как и первоначальный процесс, происходило бы в полном соответствии с законами Ньютона!

Читатель, быть может, спросит, откуда берется энергия, поднявшая стакан с пола на стол. Ответить на *этот* вопрос совсем несложно, поскольку, в то время, когда стакан *падает* со стола, энергия, которую он приобретает в процессе падения, должна куда-то *деваться*. На самом деле, энергия падающего стакана переходит в *тепло*. Атомы в осколках стакана, в воде, в ковре и половицах после удара стакана о пол будут хаотически колебаться чуть-чуть быстрее, чем до удара, т. е. осколки стакана, вода, ковер и половицы будут чуточку *горячее*, чем они были раньше (если пренебречь возможной потерей тепла за счет испарения, которое, однако, в принципе тоже обратимо). В силу

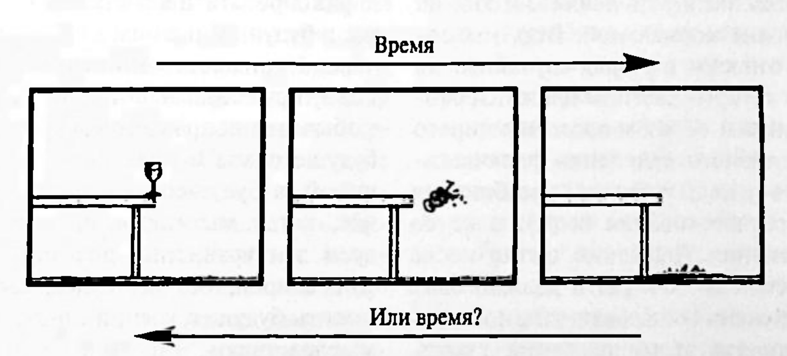


Рис. 7.2. Законы механики обратимы во времени; однако последовательность событий в направлении справа налево никогда не наблюдается, в то время как последовательность слева направо была бы вполне обычной

закона *сохранения энергии* эта тепловая энергия будет в точности равна той, которая теряется стаканом с водой при его падении со стола. Таким образом, этой маленькой порции тепловой энергии было бы как раз достаточно, чтобы поднять стакан обратно на стол! Очень важно не забыть учесть вклад тепловой энергии в общий энергетический баланс. Закон сохранения энергии, в котором учитывается также и тепловая энергия, носит название *первого начала термодинамики*. Этот закон, будучи следствием ньютоновской механики, симметричен во времени. Он не накладывает каких-либо ограничений на стакан и воду, которые бы запрещали стакану собирать себя, заполняться водой и таким вот чудесным образом запрыгивать обратно на стол.

Причина, по которой мы не наблюдаем ничего подобного в реальности, заключается в том, что «тепловое» движение атомов в осколках стекла, воде, половицах и ковре является совершенно беспорядочным, так что подавляющая часть атомов будет двигаться во всех возможных направлениях. Необходима невероятно точная координация их движений для того, чтобы восстановить стакан, вместе со всей собранной в него с пола водой, и аккуратно забросить его на стол. Можно даже утверждать, что такие слаженные движения невозможны. Точнее говоря, подобная скоординированность могла бы возникнуть только благодаря удивительной случайности, которую мы все равно отнесли бы к разряду «чудес», даже если бы она и произошла в действительности!

Однако для другого направления времени такая согласованность движений атомов является вполне нормальной. Ведь мы почему-то не относим в разряд случайных те ситуации, в которых частицы движутся скоординированным образом *после* некоторого крупномасштабного изменения физического состояния (в нашем случае — разбивания стакана и расплескивания воды), а не *до* такого изменения. Движение частиц после подобного события как раз и должно быть в высокой степени согласованным, поскольку сама природа этого движения такова, что если бы мы могли в точности обратить движение каждого отдельного атома, результирующее движение было бы именно таким, какое необходимо для восстановления, заполнения и подъема стакана в его исходное положение.

Высокая координация движения вполне приемлема и даже естественна в том случае, когда оно является *следствием* крупномасштабного изменения, а не его *причиной*. Но слова «причина» и «следствие», так или иначе, затрагивают вопрос о временной асимметрии. Используя эти термины в нашем повседневном разговорном языке, мы обычно подразумеваем, что причина должна предшествовать следствию. Но если мы пытаемся осознать физическое различие между прошлым и будущим, нам необходимо быть предельно осторожными, чтобы невольно не привнести в рассуждения наши житейские представления об этих понятиях. Я должен предупредить читателя, что избежать этого чрезвычайно трудно, но нам все же стоит попробовать. Мы должны попытаться использовать слова таким образом, чтобы они заранее не предрешали вопроса о физическом различии прошлого и будущего. В частности, если обстоятельства будут к тому располагать, нам придется иногда рассматривать причины некоторых явлений лежащими в будущем, а следствия — лежащими в прошлом! Детерминистские уравнения классической физики (или операция U в квантовой физике) никоим образом не выделяют эволюцию в направлении будущего. Они могут быть столь же хорошо применимы и для описания эволюции в прошлое. Будущее определяет прошлое точно так же, как и прошлое определяет будущее. Мы можем каким-либо образом зафиксировать некоторое состояние системы в будущем и затем использовать его для определения состояния системы в прошлом. Если, применяя наши уравнения к системе с обычным направлением времени в сторону будущего, мы можем считать прошлое причиной, а будущее — следствием, то в случае, когда мы также правомерно используем эти уравнения для описания эволюции в прошлое, мы будем вынуждены относить будущее к «причине», а прошлое — к «следствию».

Есть, однако, еще один момент, связанный с использованием терминов «причина»

и «следствие», который, на самом деле, никак не зависит от того, какие события мы относим к прошлому, а какие — к будущему. Вообразим себе гипотетическую вселенную, в которой справедливы те же симметричные во времени классические уравнения, что и в нашей вселенной, но в которой явления обычного порядка (такие, как разбивание и расплескивание стакана воды) сосуществуют с их обращениями во времени. Предположим, что наряду с обычными явлениями, стаканы воды иногда действительно собирают себя из отколовшихся кусочков, чудесным образом заполняются расплескавшейся водой и затем запрыгивают на стол; предположим также, что иногда, приготовленная яичница-болтунья снова превращается в исходный полуфабрикат, желток в ней отделяется от белка и, наконец, она запрыгивает обратно в сломанную яичную скорлупу, которая становится совершенно целой, вновь заключая в себя все свое содержимое; что кусочки сахара могут восстанавливаться из растворенного сахара в подслащенном кофе и затем самопроизвольно выпрыгивать из чашки прямо в чью-нибудь руку. Если бы мы жили в мире, в котором подобные вещи относились бы к разряду повседневных явлений, мы, очевидно, могли бы приписать «причины» таких событий не фантастической случайности, связанной с коррелированным поведением отдельных атомов, но некоторому «телеологическому воздействию», благодаря которому самовосстанавливающиеся объекты стремятся в конце концов достичь желаемой макроскопической конфигурации. «Смотрите! — могли бы воскликнуть мы. — Это повторяется. Та смесь намеревается собрать себя в другой стакан воды!» Мы, разумеется, можем принять точку зрения, согласно которой атомы направили сами себя именно так, *потому* что именно таким способом можно получить стакан воды на столе. Стакан на столе был бы в этом случае причиной, а явно беспорядочная смесь атомов на полу — «следствием» — несмотря на то, что это «следствие» теперь существует во времени раньше, чем причина. Точно также, внезапное упорядочивание движения атомов в приготовленной яичнице-болтунье не является «причиной» ее запрыгивания в целую яичную скорлупу, но есть следствие этого будущего состояния; и кусок сахара собирается и выскакивает из чашки не «потому, что» атомы движутся с такой необычайной точностью, но благодаря тому, что кто-то — находящийся в будущем — будет позднее держать этот кусок сахара в своей руке!

Конечно, мы не наблюдаем ничего подобного в нашем мире — или, лучше сказать, что мы не обнаруживаем одновременного *сосуществования* подобных вещей с явлениями обычного порядка. Ведь если бы *все*, что мы видели, было бы явлениями обратного порядка, подобного описанному выше, у нас не было бы проблем. Нам нужно было бы просто поменять местами «прошлое» и «будущее», «до» и «после» и т. д. во всех наших описаниях. Время следовало бы тогда считать текущим в направлении обратном по отношению к первоначально выбранному, и такой мир мог бы описываться так же, как и наш. Здесь я, однако, хочу рассмотреть другую возможность, в точности согласующуюся с симметричными во времени уравнениями физики, а именно — когда разбивающийся и самовосстанавливающийся стаканы могут *сосуществовать*.

В этом мире мы были бы не в состоянии восстановить привычные описания событий одним только изменением наших соглашений о направлении движения времени. Конечно, наш мир оказывается не таким — но почему? Чтобы разобраться с этим, я для начала попросил бы вас представить такой мир и подумать над тем, как описывать события, происходящие в нем. Согласитесь, что в подобном мире мы могли бы хорошо описывать крупные макроскопические конфигурации — такие как полные стаканы воды, неразбитые яйца, или кусочки сахара в руке, являющиеся «причинами»; и микроскопические, быть может, тонко скоррелированные движения отдельных атомов, представляющие «следствия» — независимо от того, лежат ли «причины» в прошлом или будущем своих «следствий».

Почему же в мире, в котором живем мы, именно причины всегда *предшествуют* следствиям или, иными словами, почему точно скоординированные движения частиц возникают только *после* крупномасштабных изменений физического состояния, а не *перед*

ними? Чтобы лучше разобраться в таком положении дел, мне нужно ввести понятие *энтропии*. Грубо говоря, энтропия системы есть мера ее явного *беспорядка*. (Позже я дам более точное определение.) Таким образом, разбитый стакан и разлитая по полу вода находятся в состоянии с большей энтропией, чем целый заполненный водой стакан на столе. Приготовленная яичница-болтунья обладает большей энтропией, чем свежее неразбитое яйцо; подслащенный кофе обладает большей энтропией, чем кофе с нерастворенным куском сахара в нем. Подобные низкоэнтропийные состояния выглядят как бы «специально упорядоченными» некоторым явным образом, а высокоэнтропийные состояния — менее «специально упорядоченными».

Здесь важно подчеркнуть, что говоря о «специальности» (или, скажем, «особенности») состояния с низкой энтропией, мы, на самом деле, имеем ввиду именно *явную* «специальность». Если этого не оговорить, то при более детальном рассмотрении мы могли бы увидеть, что высокоэнтропийные состояния в подобных ситуациях *будут* такими же «специально упорядоченными», как и низкоэнтропийные, благодаря чрезвычайно точной координации движений отдельных частиц. Например, кажущееся случайным движение молекул воды, просочившейся между половицами после того, как стакан разбился, является, на самом деле, вполне специальным: эти перемещения настолько точны, что если их *обратить*, то получится то самое исходное низкоэнтропийное состояние, в котором восстановленный стакан покоится на столе. (Это должно быть именно так, поскольку обращение всех этих движений полностью соответствует обращению направления времени, в результате которого стакан, разумеется, восстановил бы себя и запрыгнул обратно на стол.) Но подобное скоординированное движение всех молекул воды — совсем *не* та «специальность», которую мы имеем ввиду, говоря о низкой энтропии. Энтропия относится к *явному* беспорядку. Порядок же, относящийся к точной координации движений частиц, не есть явный порядок, и потому он не приводит к понижению энтропии системы. Таким образом, упорядочивание молекул разлитой жидкости, в данном случае, не учитывается, и ее энтропия остается высокой. В то же время, *явный* порядок в *восстановленном* стакане воды дает низкое значение энтропии. Все дело здесь в том, что с конфигурацией восстановленного и заполненного стакана воды совместимо относительно немного возможных движений частиц; в то время как движений, совместимых с конфигурацией слегка нагретой воды, протекающей между щелями в половицах, — существенно больше.

Второе начало термодинамики гласит, что *энтропия изолированной системы возрастает со временем* (*или остается неизменной в случае обратимых систем*). Теперь становится очевидным, что мы совершенно правильно не рассматриваем скоординированное движение частиц как признак низкой энтропии, поскольку в этом случае «энтропия» системы, в соответствии с ее определением, всегда оставалась бы постоянной. Понятие энтропии должно быть связано только с явным беспорядком. Для системы, изолированной от всей остальной вселенной, ее полная энтропия возрастает, так что, если подобная система начинает свою эволюцию из состояния с некоторой явной упорядоченностью, то с течением времени этот порядок неизбежно разрушается и присущие ей особые свойства превращаются в «бесполезно» скоординированное движение частиц.

Может показаться, что второе начало действует как некий предвестник упадка, поскольку оно утверждает существование безжалостного универсального физического принципа, напоминающего нам о том, что всякое упорядоченное состояние подвержено непрерывному разрушению. Позднее мы увидим, что это пессимистическое заключение справедливо не всегда!

Что такое энтропия?

Каково же точное определение энтропии физической системы? Мы уже знаем, что это некая мера явного беспорядка — но что означают такие не очень строгие понятия, как «явный» и «беспорядок»? Может возникнуть мысль, что энтропия — это

величина, вообще не имеющая четкого физического определения. Кроме того, имеется еще одно обстоятельство, связанное со вторым началом термодинамики, которое еще в большей степени усиливает ощущение нестрогости обсуждаемого понятия: энтропия не остается постоянной и возрастает только в так называемых *необратимых* процессах. Но что значит «необратимых»? На микроскопическом уровне, когда мы принимаем в расчет движения всех частиц, *все* процессы оказываются обратимыми! *Обычно* мы полагаем, что падение стакана со стола и его разбивание, разбалтывание яйца или растворение сахара в кофе — суть процессы необратимые; в то же время, столкновения друг с другом небольшого числа частиц — процесс обратимый, так же, впрочем, как и вообще любой процесс, в котором путем некоторых ухищрений нам удается избежать превращения кинетической энергии в тепло. Термин «необратимый» служит нам, главным образом, лишь для указания на то, что проследить за микроскопическими движениями отдельных частиц или управлять ими было невозможно. Собственно, эти неконтролируемые движения и есть «тепло». Таким образом, может создаться впечатление, будто бы понятие «необратимости» обязано своим происхождением чисто «практическим» соображениям. Мы, конечно, и в самом деле не можем *на практике* отделить белок от желтка в разболтанном яйце, хотя подобная процедура и не противоречит законам механики. Поэтому возникает вопрос: а не будет ли все-таки наше определение энтропии зависеть от того, какие процессы практически осуществимы, а какие — нет?

Как уже говорилось в главе 5, физическое понятие *энергии*, так же как и импульса, и углового момента, *имеют* вполне четкие математические определения в терминах положений частиц, их скоростей, масс и действующих на них сил. А можем ли мы сходным образом определить понятие «явного беспорядка», которое, в свою очередь, необходимо для придания точного математического смысла понятию энтропии? Очевидно, что «явное» для одного наблюдателя может не быть таковым для другого. И вообще, не находится ли это «явное» в прямой зависимости от точности, с которой тот или иной наблюдатель способен изучать данную систему? Наблюдатель, располагающий более точной измерительной аппаратурой, способен получить намного больше информации о микроскопическом строении системы, чем другой наблюдатель, использующий менее совершенное оборудование. В этом случае один наблюдатель сможет обнаружить больше «скрытого порядка», чем другой, и он, разумеется, зафиксирует более низкий уровень энтропии данной системы, чем его коллега. Может даже сложиться впечатление, что и личные эстетические вкусы каждого из наблюдателей способны оказать решающее влияние на их выбор между «порядком» или «беспорядком». Предположим, что мы пригласили некоего художника, для которого россыпь осколков стекла на полу окажется настоящим произведением «искусства упорядочивания» по сравнению с безобразным, отвратительным стаканом, банально покоящимся на краю стола! *Понизится ли* и в самом деле энтропия системы после ее оценки наблюдателем с таким тонким артистическим восприятием?

Несмотря на все проблемы, связанные с субъективностью некоторых наших суждений, понятие энтропии оказывается замечательным образом применимо всякий раз, когда речь идет о точном научном описании — каковым и является само понятие энтропии! Причина этого заключается в том, что изменения, вызванные переходами системы от порядка к беспорядку, если их выразить в терминах микроскопических положений и скоростей частиц, поистине колоссальны и (почти во всех случаях) превосходят любые заметные на глаз отличия точек зрения на то, что считать «явным порядком» на макроскопическом уровне, а что — нет. В частности, любое заключение художника или ученого, относительно того, какой из стаканов обладает бо́льшим порядком — целый или разбитый, практически не имеет никакого отношения к их реальной энтропии. Намного больший вклад в энтропию дает случайное движение частиц, вызывающее незначительное нагревание стакана и воды, и растекание воды после удара стакана с водою о пол.

Теперь, чтобы точно сформулировать понятие энтропии, вернемся к идее *фазо-*

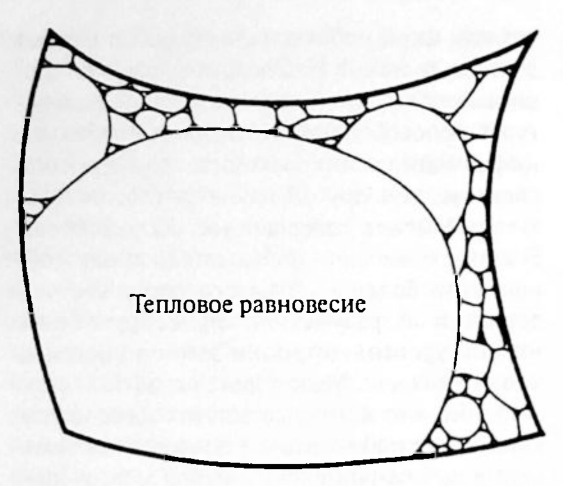


Рис. 7.3. Гранулирование фазового пространства на области, соответствующие макроскопически неотличимым состояниям. Энтропия пропорциональна логарифму фазового объема

вого пространства, введенного в главе 5. Напомним, что фазовое пространство системы имеет, как правило, гигантское число измерений, а каждая его точка изображает с максимальной детализацией мгновенную конфигурацию системы. Подчеркнем, что *«одна-единственная»* точка фазового пространства определяет одновременно положения и импульсы *всех* отдельных частиц, составляющих рассматриваемую физическую систему. Все, что нам необходимо сейчас для определения энтропии, это сгруппировать вместе все те микроскопические состояния, которые выглядят совершенно одинаковыми с точки зрения их *явных* (т. е. макроскопических) свойств. Другими словами, нам необходимо разбить наше фазовое пространство на области (рис. 7.3), в каждой из которых различные точки изображают физические системы, отличающиеся на микроскопическом уровне расположением и скоростями частиц, но которые при этом совершенно неразличимы с точки зрения макроскопического наблюдателя, для которого все точки любой такой конкретной области будут описывать *одну и ту же* физическую систему. Подобное разбиение фазового пространства на области называется *гранулированием* фазового пространства.

После такого группирования некоторые из областей могут приобрести подавляюще огромные размеры по сравнению с другими областями. Рассмотрим, к примеру, фазовое пространство газа, заключенного в ящике. Наибольшая область фазового пространства будет приходиться на состояния, в которых частицы газа практически равномерно распределены по ящику с некоторым характерным распределением скоростей, обеспечивающим однородные давление и температуру. Это характерное распределение, в некотором смысле наиболее случайное из всех возможных, называется *распределением Максвелла* — по имени Джеймса Клерка Максвелла, которого мы уже упоминали ранее. В этом случае про газ говорят, что он находится в состоянии *теплового равновесия*. Подавляющая часть точек всего фазового пространства соответствует этому тепловому равновесию, и эти точки изображают всевозможные микроскопические значения координат и скоростей отдельных частиц, которые совместимы с состоянием теплового равновесия. Эта огромная часть является, конечно, только одной из многих областей нашего фазового пространства — но она оказывается (существенно) большей всех других областей, занимая практически все фазовое пространство! Рассмотрим теперь другое возможное состояние этого газа, скажем, такое, в котором весь газ собран в одном из углов ящика. В этом случае мы будем опять иметь целое множество различных микроскопических состояний, каждое из которых описывает газ сосредоточенным в углу ящика. Все эти состояния макроскопически неразличимы, и изображающие их точки фазового пространства заполняют в нем свою область. Однако объем этой области оказывается намного меньшим объема области для состояний теплового равновесия — примерно в $10^{10^{25}}$ раз (если ящик — это метровый куб, содержащий воздух при нормальных условиях, а область в углу — сантиметровый кубик)!

Чтобы оценить различия в фазовых объемах, рассмотрим упрощенную ситуацию, в которой некоторое количество шаров распределено по большому числу ячеек. Предположим, что каждая ячейка может либо быть пустой, либо содержать один шар. Шары будут моделировать молекулы газа, а ячейки — различные положения молекул в ящике. Выделим небольшое подмножество ячеек, которое будем называть *особым*;

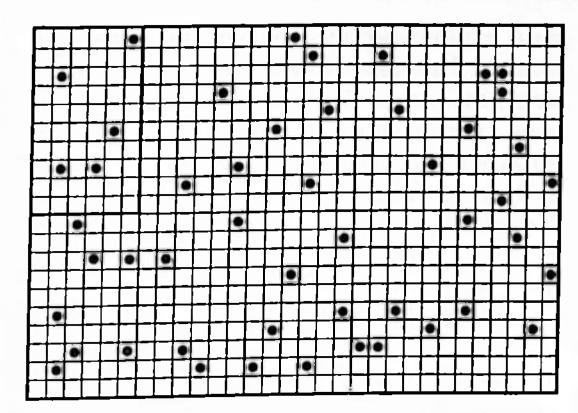


Рис. 7.4. Модель газа в ящике: некоторое количество шаров распределено по значительно большему числу ячеек. Одна десятая часть ячеек отмечены как *особые*. Эти ячейки выделены в левом верхнем углу

оно будет соответствовать положению молекул газа в углу ящика. Для определенности условимся, что ровно 1/10 часть всех ячеек особая — т. е. в случае, когда имеется n особых ячеек, не особых будет ровно $9n$ (рис. 7.4).

Мы хотим теперь случайным образом распределить m шаров среди всех ячеек и найти вероятность того, что все шары окажутся в особых ячейках. В случае, когда имеется только один шар и десять ячеек (т. е. имеется только одна особая ячейка), эта вероятность, очевидно, равна одной десятой. Тот же результат получится в случае одного шара и любого числа $10n$ ячеек (т. е. в случае n особых ячеек). Таким образом, для газа, состоящего только из *одного* атома, особая область, соответствующая «газу, собранному в углу ящика», будет иметь фазовый объем, составляющий лишь *одну десятую* всего объема «фазового пространства». Однако, если мы увеличим число шаров, вероятность того, что *все* они соберутся в особых ячейках, существенно понизится. Скажем, для *двух* шаров с двадцатью ячейками (две из которых особые) $(m = 2, n = 2)$ [1], вероятность равна 1/190; в случае ста ячеек (среди них — десять особых) $(m = 2, n = 10)$ вероятность равна 1/110; а при неограниченном увеличении числа ячеек с сохранением доли особых вероятность будет стремиться к 1/100.

Таким образом, в случае газа из двух атомов фазовый объем особой области составляет только одну *сотую* часть всего «фазового пространства». Для *трех* шаров и тридцати ячеек $(m = 3, n = 3)$, он будет составлять 1/4060 всего фазового объема, а в пределе бесконечного числа ячеек — 1/1000 — т. е. для газа из *трех* атомов объем особой части будет составлять одну *тысячную* объема всего «фазового пространства». Для четырех шаров в пределе бесконечного числа ячеек вероятность становится равной 1/10 000. Для пяти шаров — 1/100 000 и т. д. Для m шаров в пределе бесконечного числа ячеек вероятность стремится к $1/10^m$; т. е. для «газа» из m атомов фазовый объем особой области составляет только $1/10^m$ от всего «фазового объема». (Этот результат остается справедливым, если учесть также и импульсы.)

Мы можем применить теперь те же оценки к нашей ситуации с реальным газом в ящике, только в этом случае для особой области нам нужно вместо одной десятой взять одну миллионную (1/1 000 000) от общего объема ящика (т. е. отношение объемов одного кубического сантиметра и одного кубического метра). В результате, вместо значения $1/10^m$ для вероятности обнаружить все частицы газа в особой области, мы получим $1/1\,000\,000^m$, т. е. $1/10^{6m}$. Для воздуха, взятого при нормальных условиях, в нашем ящике находилось бы около 10^{25} молекул, поэтому мы принимаем $m = 10^{25}$. Таким образом, особая область фазового пространства, представляющая состояния, в которых весь газ сосредоточен в углу ящика, составляет только

$$1/10^{60\,000\,000\,000\,000\,000\,000\,000\,000}$$

часть всего фазового пространства!

Энтропия состояния — это мера *объема V* области фазового пространства, которая содержит все точки, представляющие данное состояние. Ввиду гигантской разницы между объемами, которую мы оценили выше, более удобным оказывается определять энтропию как величину, пропорциональную не самим объемам, а их *логарифмам*:

$$\text{энтропия} = k \log V.$$

[1] В общем случае n, m вероятность равна

$$\frac{C^n_m}{C^{10n}_m} = \frac{(10n)!(10n-m)!}{n!(n-m)!}.$$

Использование логарифма делает все возникающие в расчетах числа более обозримыми. Так, к примеру, логарифм[2] 10 000 000 составляет всего-навсего число, близкое к 16. Величина k — константа, называемая *постоянной Больцмана*. Ее значение приблизительно равно 10^{-23} джоулей на один градус Кельвина.

Одним из важнейших следствий использования логарифма в определении энтропии является ее *аддитивность* в случае независимых систем. Другими словами, полная энтропия двух независимых физических систем, рассматриваемых как одна система, равна *сумме* их энтропий. (Это и есть основное свойство логарифмической функции: $\log AB = \log A + \log B$. Если эти подсистемы находятся в состояниях, изображающихся областями с объемами A и B в соответствующих им фазовых пространствах, то объем фазового пространства для составной системы будет равен произведению их объемов AB, поскольку каждое микроскопическое состояние одной системы должно быть независимо учтено вместе с каждым микроскопическим состоянием другой; и, следовательно, энтропия составной системы, очевидно, будет равна именно сумме энтропий отдельных систем.)

Те гигантские отличия между размерами различных частей фазового пространства, о которых говорилось выше, в терминах энтропии будут выглядеть более скромно. Энтропия нашего кубического метра газа, как следует из предыдущих рассмотрений, оказывается всего на 1400 Дж/К ($= 14k \times 10^{25}$) больше энтропии того же газа, сосредоточенного в кубическом сантиметре «особой» области (так как $\ln(10^{6\times 10^{25}})$ составляет примерно 14×10^{25}).

Для того, чтобы определить *реальные* значения энтропии для указанных областей фазового пространства, нам осталось бы только немного позаботиться о выборе системы единиц (метры, джоули, килограммы, градусы Кельвина и т. д.). Однако, на самом деле, здесь было бы совсем неуместным заботиться об этом: для тех чудовищно огромных значений энтропии, которые я буду рассматривать в дальнейшем, выбор системы единиц не играет особой роли. Все же для определенности (и для специалистов), я скажу, что буду пользоваться так называемой *естественной системой* единиц, которая следует из законов квантовой механики и в которой постоянная Больцмана оказывается равной *единице*:

$$k = 1.$$

Второе начало в действии

Предположим, что мы привели некоторую систему в особое начальное состояние, например, поместили газ в один из углов ящика в начальный момент времени. В следующее мгновение этот газ начнет стремительно расширяться и занимать все больший и больший объем. Через некоторое время он достигнет состояния теплового равновесия. Как описывается этот процесс на языке фазового пространства? В каждый момент времени микроскопическое состояние нашего газа, зависящее от положений и скоростей всех его молекул, изображается определенной точкой фазового пространства. По мере того, как газ расширяется, эта точка как-то блуждает в фазовом пространстве, при этом точная траектория ее блужданий будет полной историей всех молекул газа. Эта точка стартует из некоторой ничтожно малой области, а именно, той, которая включает в себя всевозможные начальные микроскопические состояния, соответствующие газу, сосредоточенному в одном из углов ящика. Далее наша движущаяся точка проходит последовательность областей фазового пространства, объемы которых монотонно возрастают, что является отражением процесса расширения газа внутри ящика. По мере расширения газа, точка продолжает свое путешествие, попадая в области фазового пространства все больших и больших объемов, причем каждый новый объем будет превосходить все предшествующие по своим размерам в огромное число раз (рис. 7.5)!

[2] Используемый здесь логарифм называется натуральным, т. е. берется по основанию

$$e = 2{,}7182818285\ldots,$$

а не по основанию 10, однако это различие в нашем случае совершенно несущественно. Натуральный логарифм, $x = \ln n$, числа n — это степень, в которую мы должны возвести e, чтобы получить n, т. е. решение уравнения $e^x = n$ (см. ссылку на с. 103).

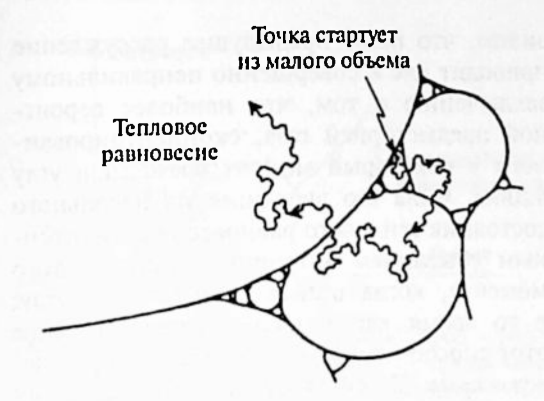


Рис. 7.5. Второе начало термодинамики в действии: с течением времени точка фазового пространства попадает в области все больших и больших объемов. Следовательно, энтропия постоянно возрастает

Всякий раз, когда точка оказывается в очередном большем объеме, у нее практически нет никаких шансов вернуться в какой-либо из предыдущих объемов меньших размеров. В конце концов, она оказывается внутри области фазового пространства наибольшего объема, соответствующей тепловому равновесию.

Этот объем занимает почти все фазовое пространство. И едва ли кто-то будет сомневаться в том, что наша точка фазового пространства в процессе своих случайных блужданий не вернется ни в какую из областей меньшего размера за любое разумное время. Можно также утверждать, что газ, достигнув состояния теплового равновесия, останется в нем практически навсегда. Мы видим, таким образом, что энтропия системы как логарифмическая мера ее фазового объема, должна так же монотонно возрастать с течением времени, как и сам фазовый объем [3].

[3] Было бы, конечно, неверным утверждать, что наша точка фазового пространства вообще никогда не достигнет ни одной из предшествующих областей меньшего объема. Если мы подождем достаточно долго, точка может снова оказаться внутри одного из них, несмотря на его ничтожно малый объем (в соответствии с *теоремой о возвращении Пуанкаре*.) Однако, в подавляющем большинстве случаев, соответствующие масштабы времен будут чудовищно велики, порядка $10^{10^{26}}$ лет, в случае газа, собравшегося в сантиметровом кубике в одном из углов ящика. Это на много порядков больше времени существования вселенной. Я не собираюсь обсуждать эту возможность в дальнейшем из-за ее практической нереализуемости.

Может показаться, что, наконец-то, мы обрели *ключ к пониманию* второго начала термодинамики! В самом деле, мы можем предположить, что наша точка фазового пространства движется совершенно хаотически, и, стартуя из некоторого крохотного объема фазового пространства, соответствующего *малому* значению энтропии, будет в дальнейшем с большой вероятностью попадать внутрь все больших и больших объемов, соответствующих все возрастающим значениям энтропии.

Есть, однако, нечто странное в том выводе, к которому, похоже, мы пришли путем такого рассуждения. Похоже, мы пришли к выводу с явной *асимметрией во времени*. Если энтропия *возрастает* в *прямом* направлении времени, то, следовательно она должна *убывать* в *обратном* направлении. Но откуда взялась эта временна́я асимметрия? Мы абсолютно уверены в том, что не использовали в наших рассуждениях никаких несимметричных во времени законов и соображений. Эта временна́я асимметрия, на самом деле, является прямым следствием того обстоятельства, что наша система *начала эволюционировать* из особого (низкоэнтропийного) состояния, и наше наблюдение за ее *последующей* эволюцией выявило факт возрастания ее энтропии. Такое возрастание, конечно же, находится в полном соответствии с поведением систем в нашей реальной вселенной. Но мы могли бы с равным успехом применить те же самые рассуждения и для обратного направления времени. Именно, мы могли бы опять создать некоторое низкоэнтропийное состояние в начальный момент времени, но теперь задаться вопросом: какова наиболее вероятная последовательность состояний, *предшествующих* этому начальному состоянию?

Попробуем теперь порассуждать в таком обратном направлении. Как и ранее, выберем в качестве низкоэнтропийного состояния газ, сосредоточенный в одном из углов ящика. В этом случае наша точка фазового пространства будет в начальный момент времени находиться в той же ничтожно малой области фазового пространства, что и ранее. Но теперь мы попробуем проследить за ее *предыдущей* историей. Если мы

представим, что эта точка, также как и ранее, движется совершенно хаотично, мы обнаружим, по мере наблюдения за последовательностью ее прошлых состояний, что сначала она достигает того же значительно большего объема фазового пространства, что и ранее, соответствующего некоторой промежуточной стадии расширения не в состоянии теплового равновесия. Затем, проходя через последовательность областей с монотонно растущими и сильно отличающимися друг от друга объемами, в самом удаленном прошлом она попадает в тот самый наибольший объем, соответствующий тепловому равновесию. *Теперь* мы, очевидно, приходим к следующему наиболее вероятному сценарию предшествующей истории газа, сосредоточенного в некоторый момент времени в одном из углов ящика: находясь в состоянии теплового равновесия, газ начинает все больше и больше концентрироваться в направлении одного из углов ящика и, наконец, весь собирается в небольшом объеме в этом углу. Во время подобного процесса энтропия должна была бы *убывать*: ее начальное значение в тепловом равновесии велико, затем оно непрерывно падает до тех пор, пока не достигнет очень низких значений, соответствующих газу, собранному в небольшом объеме в углу ящика.

Все это, конечно, имеет совсем мало общего с тем, что происходит в действительности в нашей вселенной! Энтропия никогда не убывает подобным образом; она *возрастает*. Если бы в некоторый момент времени газ действительно был бы сконцентрирован в одном из углов ящика, то, скорее всего, *ранее*, газ надежно удерживался в этом углу перегородкой, которую затем внезапно убрали. А может быть, газ удерживался там самопроизвольно, будучи охлажденным до температуры его твердого или жидкого состояния, а затем был очень быстро разогрет и, в результате, перешел в газообразную фазу. В любом случае, энтропия этих предшествующих состояний была бы даже еще *ниже*, чем исходного. Второе начало, несомненно, оставалось бы справедливым и в этих случаях, и энтропия бы все время возрастала — т. е. при *обратном* течении времени она бы, как нетрудно понять, *убывала*. *Теперь* мы отчетливо видим, что наше предыдущее рассуждение приводит нас к совершенно неправильному заключению о том, что наиболее вероятной предысторией газа, сконцентрированного в некоторый момент времени в углу ящика, была его эволюция из начального состояния теплового равновесия с монотонным убыванием энтропии вплоть до того момента, когда весь газ собрался в углу; в то время как в нашем реальном мире этот способ оказывается чрезвычайно *маловероятным*. В действительности, газ должен был начинать свою эволюцию из состояния с гораздо *меньшим* значением энтропии и энтропия должна была монотонно *возрастать*, проходя через все свои промежуточные значения вплоть до момента времени, когда весь газ соберется в углу.

Таким образом, наши рассуждения, опирающиеся на свойства случайных блужданий точки в фазовом пространстве, оказываются вполне удовлетворительными, когда мы применяем их для предсказания будущей эволюции системы и совершенно неудовлетворительными для восстановления ее прошлой эволюции. Именно, мы получаем, что наиболее вероятным *будущим* газа, который начинает эволюционировать из угла ящика, будет его конечное состояние теплового равновесия, а не внезапное появление перегородки или внезапное замерзание или сжижение газа. Столь странные сценарии будущего как раз и могли бы послужить примерами процессов, протекающих с понижением энтропии, которые совершенно исключаются нашей трактовкой процессов в фазовом пространстве. Но в направлении *прошлого*, именно такие «странные» сценарии и могли бы иметь место и, более того, они совсем не выглядят странными. Наши рассуждения, связанные с представлением процессов в фазовом пространстве, дали нам совершенно неправильный ответ при попытке применить их к обратному направлению времени!

Очевидно, все это бросает тень сомнения на наши исходные рассуждения. Получается, что мы *не* обрели никакого ключа к пониманию второго начала. Единственный достоверный вывод, который мы можем сделать из наших рассуждений, заключается в следующем: если фиксировано какое-



Рис. 7.6. Если мы интерпретируем ситуацию, изображенную на рис. 7.5 в обратном направлении времени, мы «восстановим» такое прошлое, в котором энтропия должна возрастать от ее настоящего значения. Это катастрофически противоречит наблюдениям

либо начальное низкоэнтропийное состояние (скажем, газ, собранный в углу ящика), то *в отсутствии каких-либо факторов, ограничивающих систему*, следует ожидать возрастания энтропии в *обоих* направлениях времени по отношению к энтропии данного состояния (рис. 7.6). Это утверждение не сработало в нашем случае в направлении прошлого именно из-за того, что подобные ограничения *имелись*. Безусловно существовало нечто, ограничивающее систему в прошлом. Это было что-то такое, что просто *вынудило* энтропию быть низкой в прошлом. Таким образом, стремление энтропии к возрастанию в будущем совсем неудивительно. Высокоэнтропийные состояния, в некотором смысле — состояния «естественные», которые не требуют какого-либо объяснения причин своего существования. Настоящей загадкой являются низкоэнтропийные состояния в прошлом. А что ограничивало наш мир и сделало его энтропию в прошлом столь низкой? Именно повсеместное присутствие состояний с ничтожно малой энтропией и есть самый удивительный факт той действительной вселенной, в которой мы живем, хотя такие состояния настолько привычны для нас, что мы, как правило, перестаем им удивляться. Мы сами представляем собой системы с пренебрежительно малой энтропией. Все вышеизложенные соображения подводят нас к мысли о том, что мы можем легко объяснить стремление энтропии увеличиваться с течением времени для системы, начинающей эволюцию из некоторого *заданного* низкоэнтропийного состояния. Но что действительно *достойно* удивления, так это тот факт, что энтропия оказывается монотонно убывающей по мере того, как мы продолжаем ее измерять во все более и более отдаленном прошлом этой системы!

Источник низкой энтропии во Вселенной

Теперь мы попытаемся понять, откуда же все-таки берется такая «удивительно» низкая энтропия в том реальном мире, где мы живем. И начнем мы, в первую очередь, с самих себя. Если мы сумеем разобраться с вопросом о природе нашей собственной низкой энтропии, то, наверное, сумеем найти ее источник и для газа, удерживаемого перегородкой, и для стакана воды на столе, и для яйца над шкворчащей сковородой, и для кусочка сахара над чашкой кофе. В каждом из перечисленных случаев прямо или косвенно в дело были замешаны или одно лицо, или группа людей (и даже курица!). Создание подобных низкоэнтропийных состояний в значительной мере было связано с использованием некоторой небольшой части нашей собственной низкой энтропии. Но это, возможно, была не единственная причина. Не исключено, что для откачки газа за перегородку в углу ящика использовался специальный вакуумный насос.

Если насос был не ручной, то, наверное, для получения низкоэнтропийной энергии, необходимой для этого процесса, было использовано какое-нибудь «природное топливо» (например, нефть). Возможно также, что насос имел электрический привод и, в некоторой степени, использовал низкоэнтропийную энергию, заключенную в урановом топливе атомной энергетической станции. Я вернусь ко всем этим внешним низкоэнтропийным источникам позже, но сперва давайте разберемся с низкой энтропией в нас самих.

Откуда же и в самом деле берется наша собственная столь малая энтропия? Строительный материал для наших тел — это продукты, которые мы едим, и кислород, которым мы дышим. Существует довольно расхожее мнение, что продукты и кислород необходимы нам лишь для получения

энергии, но, на самом деле, это верно лишь отчасти. Потребляемые нами продукты действительно окисляются кислородом, который мы вдыхаем, и это обеспечивает нас энергией. Но большая часть этой энергии снова покидает наши тела, главным образом, в виде тепла. Поскольку энергия сохраняется, и поскольку реальное энергетическое содержание наших тел остается более или менее неизменным на протяжении всей нашей взрослой жизни, то нет никакой необходимости и увеличивать его. Нам вполне достаточно той энергии, которая содержится в наших телах в настоящий момент. Иногда мы, действительно, увеличиваем собственное энергетическое содержание, когда наращиваем вес — но это, как правило, совсем нежелательно! Также, начиная с детского возраста, по мере взросления и роста нашего тела, мы значительно увеличиваем свое энергетическое содержание; но речь сейчас идет совсем не об этом. Вопрос заключается в том, как нам удается поддерживать свою жизнь на всем ее протяжении (в основном во взрослый период). Для этого нам совсем не требуется увеличивать свое энергетическое содержание.

Тем не менее, нам действительно необходимо пополнять энергию, которую мы постоянно теряем в виде тепла. Несомненно, что чем более мы «энергичны», тем большее количество энергии мы теряем таким образом. Вся эта энергия должна быть восстановлена. Тепло — это самая неупорядоченная, т. е. самая высокоэнтропийная форма энергии в ряду остальных. Мы потребляем энергию в низкоэнтропийной форме (продукты и кислород), а выделяем ее в форме высокоэнтропийной (тепло, углекислый газ, экскременты). Нам не нужно как-то вылавливать энергию из окружающей среды, так как энергия *сохраняется*. Но мы непрерывно боремся со вторым началом термодинамики. Энтропия не постоянна — она все время растет. Для поддержания нашей жизни нам необходимо сохранять тот низкий уровень энтропии, который имеется внутри нас. Это нам удается благодаря потреблению низкоэнтропийной комбинации продуктов и атмосферного кислорода, их взаимодействию в наших телах и выделению энергии, которую иначе мы бы усвоили, в высокоэнтропийной форме. Таким образом, мы можем предохранять энтропию наших тел от возрастания и можем поддерживать (и даже совершенствовать) свою внутреннюю организацию (см. Шредингер [1967]).

А откуда берется этот запас низкой энтропии? Если речь идет о мясе (или грибах!), то эти продукты, как и мы сами, должны были использовать внешние низкоэнтропийные источники следующего уровня, для обеспечения и поддержания своей низкоэнтропийной структуры. Это только переводит вопрос об источнике внешней низкой энтропии на что-то еще. Предположим теперь, что мы (или животные, или грибы) потребляем растения. Все мы, на самом деле, должны быть чрезвычайно благодарны зеленым растениям — прямо или косвенно — за их замечательную способность потреблять атмосферный углекислый газ, разделять углерод и кислород и использовать углерод в качестве строительного материала для своих организмов.

Этот процесс, называемый фотосинтезом, приводит к сильному понижению энтропии. Мы сами используем это низкоэнтропийное разделение, в конечном счете, просто соединяя снова кислород и углерод внутри наших тел. Каким же образом зеленые растения совершают подобное чудо? Они используют солнечный свет. Этот свет переносит энергию с Солнца на Землю в сравнительно низкоэнтропийной форме — в виде фотонов видимого света. Земля, включая и ее обитателей, не задерживает эту энергию надолго, а переизлучает ее целиком обратно в окружающее пространство. Однако эта переизлученная энергия находится уже в высокоэнтропийной форме, а именно, в виде так называемого «радиационного тепла», т. е. инфракрасных фотонов. В противоположность общепринятому мнению, Земля вместе с ее обитателями не получает энергии от Солнца! Вся роль Земли здесь сводится к тому, чтобы принять энергию в низкоэнтропийной форме, а затем рассеять ее обратно в окружающее пространство, но уже как энергию с высокой энтропией (рис. 7.7). Таким образом, Солнце служит для нас мощным источником низкой энтропии. Мы (благодаря упомянутой замечательной способности растений)



Рис. 7.7. Так мы используем Солнце — раскаленный шар среди темноты космического пространства

это используем, выделяя некоторую небольшую ее часть и преобразуя ее в удивительные по своей сложности структуры наших организмов.

Давайте теперь в общих чертах рассмотрим, что происходит с энергией и энтропией относительно Солнца и Земли. Солнце излучает энергию в виде фотонов видимого диапазона длин волн. Часть из них поглощается Землей, а затем переизлучается в виде фотонов инфракрасного диапазона. Решающее значение здесь имеет тот факт, что видимые фотоны имеют большую частоту, чем инфракрасные и, следовательно, большую энергию, приходящуюся на одну частицу. (Вспомните формулу Планка $E = h\nu$, приведенную на с. 211. Она как раз и говорит о том, что энергия фотона пропорциональна его частоте.) Так как одиночный видимый фотон обладает большей энергией, чем одиночный инфракрасный, то видимых фотонов, падающих на Землю, должно быть меньше, чем инфракрасных, испускаемых Землей, причем ровно настолько, чтобы соблюдался баланс между падающей и излученной энергиями. А значит, энергия, переизлучаемая Землей в окружающее пространство, распределяется по гораздо большему числу степеней свободы, чем энергия, получаемая Землей от Солнца. Из-за этого большого числа задействованных степеней свободы соответствующий объем в фазовом пространстве электромагнитного поля также оказывается значительно большим у переизлученных фотонов по сравнению с фазовым объемом падающих и, следовательно, энтропия системы фотонов после переизлучения существенно возрастает. Зеленые растения, потребляя энергию в низкоэнтропийной форме (сравнительно небольшого числа видимых фотонов) и переизлучая ее в высокоэнтропийной форме (сравнительно большого числа инфракрасных фотонов), одновременно обеспечивают себя необходимой низкой энтропией, а нас — жизненно необходимым разделением углерода и кислорода.

И все это возможно благодаря тому, что Солнце — это горячее пятно на небе! Дело в том, что небо находится в термодинамически неравновесном состоянии: один его небольшой участок, а именно, тот, который и занимает Солнце, имеет температуру, намного превышающую температуру оставшейся его части. Благодаря этому мы и оказываемся обеспечены мощным источником низкой энтропии. Земля получает энергию от этого горячего пятна в низкоэнтропийной форме (немного фотонов) и переизлучает ее в холодные области неба в высокоэнтропийной форме (много фотонов).

А почему Солнце является этим горячим пятном? Каким образом оно приобрело столь высокую температуру и затем смогло поддерживать низкоэнтропийные состояния других систем? Ответ заключается в том, что изначально оно образовалось из однородного газового облака (главным образом — водорода) посредством гравитационного сжатия. В ходе этого процесса, еще на ранних стадиях своего образования, Солнце разогрелось. Оно продолжало бы и далее сжиматься и разогреваться, если бы, при некоторых определенных давлении и температуре, в игру не вступил другой источник энергии негравитационной природы, а именно, термоядерные реакции: слияние ядер водорода в ядра гелия с выделением энергии. Без термоядерных реакций Солнце было бы намного горячее и меньше, чем сейчас, оставаясь таким до самого момента своей звездной смерти. Термоядерные реакции не дали Солнцу стать слишком горячим, приостановив его дальнейшее сжатие и стабилизировав температуру Солнца на том уровне, который оказался вполне пригоден для нашей жизни, одновременно продлив при этом период его свечения.

Важно отметить, однако, что хотя термоядерные реакции и играют очень важную роль в происхождении и установлении количественных характеристик солнечной энергии, именно гравитация является здесь решающим фактором. (На самом деле, возможность термоядерных реакций дает существенный вклад в низкую энтропию Солнца, но учесть энтропию, обусловленную слиянием ядер весьма непросто, и детальное обсуждение этого вопроса только усложнило бы наши рассуждения, не изменяя окончательного вывода.) [3] Без гравитации Солнце вообще не могло бы существовать! Оно продолжало бы светить и без термоядерных реакций (хотя в этом случае его излучение было бы губительным для нас), но без гравитации оно не светило бы вообще, поскольку именно гравитационное взаимодействие связывает вещество Солнца и обеспечивает необходимые температуру и давление. Без гравитации вместо Солнца мы имели бы холодный и рассеянный газ — такой же «мертвый», как и остальное космическое пространство вокруг нас.

Нам осталось обсудить вопрос об источнике низкой энтропии различных видов «природного топлива» на Земле; но суть и в этом случае остается прежней. В соответствии с общепринятыми взглядами, вся нефть (и природный газ) образовались из доисторической растительности. И снова растения оказываются источником низкой энтропии. Поскольку доисторическая растительность имела благодаря Солнцу низкую энтропию, то мы опять возвращаемся к гравитации, которая формирует Солнце из рассеянного газа. Существует интересная «альтернативная» теория происхождения нефти на Земле, выдвинутая Томасом Голдом, который оспаривает традиционный подход, утверждая, что доисторическая растительность не могла послужить источником такой гигантской массы гидрокарбонатов на Земле. Голд полагает, что нефть и природный газ были захвачены внутренностью Земли во время ее формирования, и с тех пор они непрерывно просачиваются наружу, накапливаясь в подземных пустотах и по сей день [4]. Согласно теории Голда, синтез нефти в любом случае должен был происходить под действием солнечного света, хотя на этот раз в космосе, прежде чем сформировалась Земля. Но и здесь за все отвечает Солнце, которое сформировала гравитация.

А что можно сказать по поводу низкоэнтропийной ядерной энергии изотопа урана-235, который используется в ядерных реакторах? Она имеет своим источником не само Солнце (хотя вполне и могла быть связана с Солнцем на некоторой стадии), а какие-то другие звезды, которые взорвались много миллиардов лет назад во время вспышек сверхновых. В действительности, этот материал образовался в результате большого числа таких вспышек. Он рассеялся в пространстве после взрыва, часть его случайно соединилась (под воздействием Солнца) и обеспечила Землю тяжелыми элементами, включая и весь запас урана-235 на ней. Каждое ядро, с его низкоэнтропийным запасом энергии, возникло в результате грандиозного ядерного процесса, происходившего во время вспышки сверхновой. Этот взрыв, в свою очередь, был следствием гравитационного коллапса [5] звезды, которая была слишком массивна, чтобы сдерживать этот коллапс одними только силами теплового давления. После такого коллапса и последующего взрыва обычно остается только небольшое ядро — возможно, в виде так называемой нейтронной звезды (подробнее о них чуть позже!). Эта звезда должна была получиться в результате гравитационного сжатия рассеянного газового облака, и большая часть ее исходного вещества — включая и наш уран-235 — должна была быть выброшена обратно в космическое пространство. При этом, однако, благодаря гравитационному сжатию, в целом произошел колоссальный выигрыш в энтропии, заключенной в ядре оставшейся нейтронной звезды. И снова именно гравитация окончательно все расставила по местам, конденсируя (на последних этапах — стремительно) рассеянный газ в нейтронную звезду.

Таким образом напрашивается вывод, что вся та удивительно низкая энтропия, которую мы обнаруживаем вокруг себя — и которая составляет наиболее загадочную сторону второго начала термодинамики — должна быть приписана тому, что огромный выигрыш в энтропии может быть получен в процессе гравитационного сжатия

рассеянного газа в звезды. А откуда взялся весь этот рассеянный газ? Здесь для нас важно, что в самом начале этот газ был *рассеянным*, благодаря чему человечество было обеспечено огромным запасом низкой энтропии, которого нам хватало до сих пор и хватит еще на продолжительный период в будущем.

Именно возможность собирания этого газа в гравитационные сгустки и дала нам второе начало термодинамики. Более того, эти сгустки не просто послужили основанием второго начала, но дали нечто намного более точное и определенное, чем простое утверждение: «Энтропия мира вначале была очень низкой». Ведь энтропия могла быть дана нам низкой и многими другими способами, например, в ранней вселенной мог бы иметь место космологический «явный порядок» совсем другого рода, чем тот, с которым мы сталкиваемся в действительности.

(Представьте себе, что ранняя вселенная была бы правильным додекаэдром — как это могло видеться Платону — или имела бы какую-нибудь другую самую невероятную геометрическую форму. Это был бы, конечно, самый настоящий «явный порядок», но совсем не тот, который мы ожидали бы обнаружить в действительной ранней вселенной!) Мы должны разобраться в том, откуда взялся весь этот рассеянный газ, для чего нам необходимо обратиться к существующим космологическим теориям.

Космология и Большой взрыв

Наша Вселенная на всех масштабах, доступных для наблюдений с помощью самых мощных оптических и радиотелескопов, оказывается в целом довольно однородной; и, что еще более впечатляет, она расширяется. При этом, чем большее расстояние разделяет нас и удаленные объекты — галактики (или совсем далекие квазары), тем с большей скоростью эти объекты удаляются от нас. Все выглядит так, как будто сама Вселенная родилась в результате гигантского взрыва, который принято называть *Большим взрывом*, имевшим место несколько десятков миллиардов лет назад[4]. Убедительным свидетельством в пользу однородности Вселенной и существования Большого взрыва оказалось открытие *чернотельного фонового излучения*. Это тепловое излучение, состоящее из фотонов, не имеющих явного источника и движущихся совершенно хаотично, имеет температуру 2,7° по абсолютной шкале (2,7 K), т. е. −270,3° Цельсия или 454,4° ниже нуля по Фаренгейту. И хотя кажется, что эта температура очень низка (а так оно, в действительности, и есть!), это излучение представляет собой остаток вспышки самого́ Большого взрыва! Из-за колоссального расширения, которое испытала Вселенная с момента Большого взрыва, начальный пылающий сгусток вещества распределился впоследствии по гигантскому объему. Температура Большого взрыва намного превышала все мыслимые значения, с которыми мы имеем дело, но из-за расширения она понизилась до той совершенно ничтожной величины, которую чернотельное фоновое излучение имеет сегодня.

Впервые существование фонового излучения было теоретически *предсказано* американским физиком и астрономом русского происхождения Георгием Гамовым в 1948 году, на основе общепринятой ныне теории Большого взрыва. А в 1965 году Пензиас и Вильсон впервые (и совершенно случайно) обнаружили его.

Я собираюсь задать вопрос, который обычно многих озадачивает. Если все далекие галактики во Вселенной удаляются от нас, не означает ли это, что мы сами занимаем какое-то особое центральное положение во Вселенной? Оказывается, нет! Точно такое же разбегание наблюдалось бы и из *любого другого* места во Вселенной. В больших масштабах расширение Вселенной однородно и все положения во Вселенной совершенно равноправны.

Часто это положение иллюстрируют с помощью надуваемого шара (рис. 7.8). Пусть

[4] В настоящее время эта цифра уточняется. Современные оценки возраста Вселенной колеблются между 6×10^9 и $1{,}5 \times 10^{10}$ лет. В любом случае эти цифры намного превосходят те 10^9 лет, которые полагались в качестве оценки возраста Вселенной сразу после открытия ее расширения Эдвином Хабблом приблизительно в 1930 году.

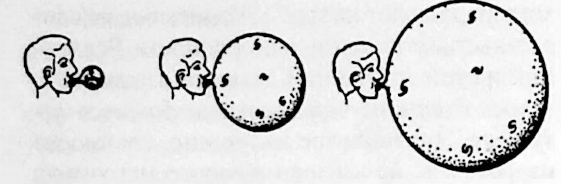


Рис. 7.8. Расширяющаяся вселенная очень напоминает поверхность надуваемого шара. Все галактики удаляются друг от друга

пятнышки на шаре изображают различные галактики, а сама двумерная поверхность шара — все трехмерное пространство вселенной. Ясно, что относительно *произвольно выбранной* точки на шаре *все* остальные точки удаляются. В этом смысле все точки шара равноправны. Точно так же, наблюдая из любой выбранной нами галактики, мы обнаружим изотропное удаление всех остальных галактик.

Раздувающийся шар дает хорошее представление об одной из трех общепринятых моделей вселенной, называемых моделями *Фридмана—Робертсона—Уокера* (ФРУ), а именно: пространственно замкнутой ФРУ-модели с *положительной кривизной*. В двух других ФРУ-моделях (с нулевой и отрицательной кривизной) вселенная расширяется подобным же образом, но вместо пространства конечного объема, которое изображает шар, мы имеем *бесконечную* вселенную с бесчисленным множеством галактик.

Из этих двух моделей наиболее проста для понимания модель с *евклидовой* пространственной геометрией, т. е. с *нулевой* кривизной. Будем изображать всю пространственную вселенную обычной плоскостью, на которой помечены точки, изображающие галактики. По мере эволюции вселенной во времени эти галактики одинаковым образом удаляются друг от друга. Попробуем представить развитие этого процесса в *пространстве-времени*. Там мы будем иметь совокупность различных «мгновенных» евклидовых плоскостей, сложенных в стопку, которая изображает всю вселенную сразу во всей ее пространственно-временной целостности (рис. 7.9). Галактики теперь будут иметь вид некоторых *кривых*, называемых *мировыми линиями* историй галактик, и эти кривые будут расходиться друг от друга в направлении будущего. И снова

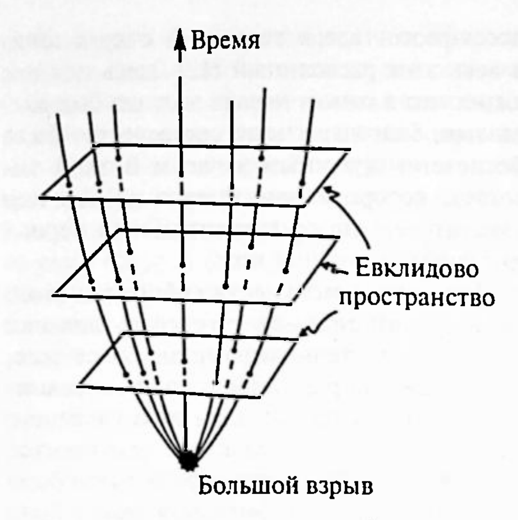


Рис. 7.9. Пространственно-временная картина расширяющейся вселенной с евклидовыми пространственными сечениями (показаны только два пространственных измерения)

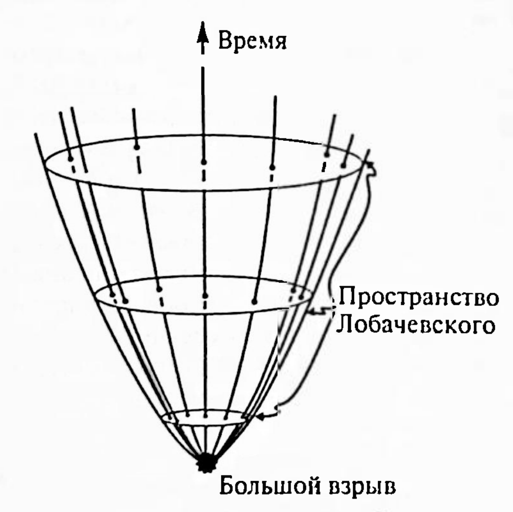


Рис. 7.10. Пространственно-временная картина расширяющейся вселенной с пространственными сечениями Лобачевского (показаны только два пространственных измерения)

все мировые линии галактик оказываются равноправными.

В оставшейся ФРУ-модели с *отрицательной* кривизной в качестве пространственной геометрии берется *неевклидова* геометрия *Лобачевского*, которая подробно описана в главе 5 и проиллюстрирована картиной Эшера (рис. 5.2, с. 154). Для построения полной пространственно-временной картины нам необходимо все «мгновенные» пространства Лобачевского расположить вплот-

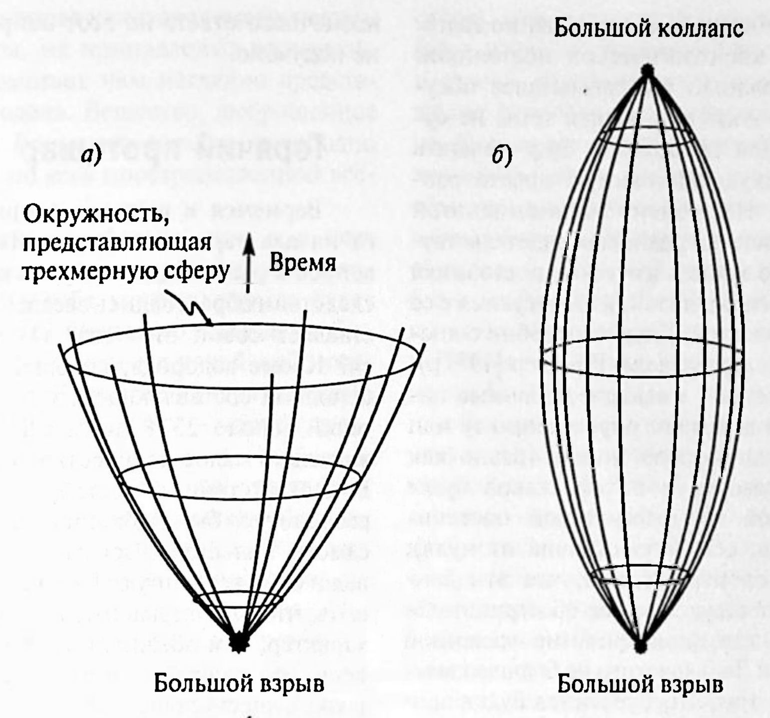


Рис. 7.11. *а*) Пространственно-временная картина расширяющейся вселенной со сферическими пространственными сечениями (показано только одно пространственное измерение); *б*) На конечной стадии вселенная испытывает большой коллапс

ную одно над другим в порядке их следования (рис. 7.10) [6]. Мировые линии галактик будут опять изображаться расходящимися в направлении будущего кривыми, причем все галактики и здесь оказываются совершенно равноправными.

Конечно, при таком описании мы для большей наглядности изображаем не все четыре измерения, а показываем лишь трехмерное пространственно-временное сечение, убирая одно измерение (точно также, как мы это делали в главе 5, с. 183). Но даже и этого оказывается недостаточно, чтобы наглядно изобразить пространство-время положительной кривизны — необходимо убрать еще одно измерение! Сделаем это и изобразим замкнутое трехмерное пространство вселенной положительной кривизны (одномерной) *окружностью*, а не (двумерной) сферой, которой была поверхность шара. По мере расширения вселенной размер этой окружности растет и мы можем изобразить все пространство-время, накладывая одну окружность на другую (каждую — для своего момента времени) и получая в результате искривленный конус (рис. 7.11 *а*).

Из уравнений Эйнштейна общей теории относительности следует, что такая замкнутая вселенная не может расширяться вечно. После того, как ее размер достигнет некоторого максимального, она начнет сжиматься и, в конце концов, сколлапсирует в точку, испытав при этом как бы большой взрыв наоборот (рис. 7.11 *б*). Этот большой взрыв наоборот иногда называют *большим коллапсом*. Во ФРУ-моделях с отрицательной и нулевой кривизной вселенная уже не коллапсирует повторно. Вместо большого коллапса, она продолжает неограниченно расширяться.

Так, во всяком случае, обстоит дело в *стандартной* общей теории относительности, в которой так называемая *космологическая постоянная* полагается равной нулю. Подбирая ненулевое значение этой космологической постоянной, можно получить или неограниченную вселенную, испытывающую большой коллапс, или конечную вселенную положительной кривизны, кото-

рая будет расширяться неопределенно долго. Присутствие космологической постоянной немного усложнило бы дальнейшее обсуждение, но в контексте нашей темы не существенно. Для простоты я буду полагать космологическую постоянную просто равной нулю[5]. На момент написания этой книги эмпирические данные свидетельствуют о том, что космологическая постоянная должна быть очень малой, и согласуются с ее нулевым значением. (Более подробно о космологических моделях см. Риндлер [1977].)

К сожалению, имеющиеся данные наблюдений не выделяют определенно ту или иную космологическую модель (равно как ничего не говорят и о том, каков будет эффект малой космологической постоянной в случае, если она отлична от нуля). С другой стороны, кажется, что эти данные свидетельствуют скорее об отрицательной пространственной кривизне вселенной (с геометрией Лобачевского на больших масштабах); и о том, что Вселенная будет продолжать расширяться неограниченно долго. Основанием для такого вывода служит, главным образом, то количество видимого вещества во вселенной, которое доступно непосредственным наблюдениям. Однако в пространстве может оказаться рассеянным и огромное количество невидимой материи, в случае чего вселенная будет обладать положительной кривизной и может в конце своей эволюции испытать большой коллапс, хотя это произойдет за промежуток времени, намного превосходящий 10^{10} лет — время существования Вселенной. Чтобы такой коллапс стал возможным, распределенного по пространству невидимого вещества — так называемой «темной материи» — должно быть раз в тридцать больше того количества видимой материи, которую мы наблюдаем в телескопы. Имеются надежные косвенные свидетельства в пользу того, что значительное количество темной материи все же присутствует, но вот достаточно ли ее, чтобы гравитационно замкнуть вселенную (или хотя бы сделать ее плоской) — привести ее к коллапсу — однозначного ответа на этот вопрос пока еще не получено.

[5] Эйнштейн ввел в теорию космологическую постоянную в 1917 году, но впоследствии, в 1931 году, отказался от нее, говоря о ней как о своей «самой большой ошибке».

Горячий протошар

Вернемся к вопросу о природе второго начала термодинамики. Мы свели этот вопрос к рассеянному газу, из которого впоследствии образовались звезды. Но что представляет собой этот газ? Откуда он взялся? Кроме водорода, который является его основной составляющей, туда входит также гелий (около 23 % по массе) и пренебрежительно малое количество других веществ. В соответствии с общепринятой теорией, весь этот газ был выброшен в результате того самого Большого взрыва, который образовал и саму вселенную. Важно, однако, понимать, что этот взрыв имел совершенно иной характер, чем обычный взрыв, при котором вещество выбрасывается из его эпицентра в уже существующее окружающее пространство. В нашем случае само пространство *возникает* в результате взрыва и никакого эпицентра нет (или не было) вообще! Такую ситуацию проше всего представить себе в случае пространства положительной кривизны. Обратимся снова к рис. 7.11 или к рис. 7.8 с раздувающимся шаром. Никакого «предсуществующего пустого пространства», в которое могло бы извергаться вещество, порожденное взрывом, нет. Само пространство, т. е. «поверхность шара» возникает в результате взрыва. Надо отдавать себе отчет в том, что только из соображений наглядности мы изобразили на рисунках (для случая положительной кривизны) «объемлющие пространства» — Евклидово пространство, в котором находится шар на рис. 7.8, и трехмерное пространство, в котором изображено пространство-время на рис. 7.11. Ни одно из этих объемлющих пространств не имеет какого-либо физического смысла. Пространство снаружи и внутри шара всего лишь помогает нам наглядно представить его поверхность. Именно поверхность шара *и только она* представляет физическое пространство вселенной. Совершенно очевидно, что нет никакого центра, из которого бы извергался материал вселенной в процессе Большого взрыва.

Точка, которая кажется геометрическим центром шара, не принадлежит вселенной, она лишь помогает нам наглядно представить нашу модель. Вещество, выброшенное в результате Большого взрыва, однородно рассеивается по *всей* пространственной вселенной.

Точно так же обстоит дело и в двух других стандартных моделях (хотя наглядно представить себе картину взрыва будет немного труднее). Вещество никогда не было сконцентрированным в какой-либо точке пространства. Напротив, оно равномерно заполняло *все* пространство — причем, с самого начала.

Такая картина лежит в основе теории *горячего большого взрыва*, называемой *стандартной моделью*. Согласно этой теории, вселенная, сразу после своего возникновения, была чрезвычайно разогретой и находилась в состоянии *горячего протошара*. В результате довольно кропотливых вычислений, мы имеем некоторое представление о природе и начальном составе этого шара (т. е. самой ранней вселенной), а также о том, как менялся этот состав по мере расширения и остывания протошара. Достоин удивления тот факт, что для описания вселенной, находящейся в столь отличном от нынешнего состоянии, вообще оказались возможными какие-либо правдоподобные вычисления. Правда, физические принципы, на которых эти вычисления основаны, работают, пока мы не интересуемся событиями, происходившими во вселенной *в первые* десятитысячные доли секунды после Большого взрыва! Начиная с этого момента и на протяжении последующих трех минут после Большого взрыва поведение вселенной изучено в деталях (см. Вайнберг [1977]) и, что удивительно, современные физические теории, основанные на экспериментальных наблюдениях нынешней вселенной, которая очень сильно отличается по свойствам от той, далекой ранней вселенной, оказываются вполне пригодными для ее описания [7]. Из этих вычислений следует, что во вселенной в однородном рассеянном состоянии должно было находиться большое количество фотонов (т. е. свет), электронов и протонов (две составные части водорода), небольшое количество α-частиц (ядра гелия), еще меньшее количество дейтронов (ядра дейтерия, тяжелого изотопа водорода) и совсем незначительное количество ядер других элементов; а также, вполне вероятно, большое количество всевозможных «невидимых» частиц, которые весьма неохотно обнаруживают себя наблюдателю. Эти *материальные* составляющие вселенной (главным образом, протоны и электроны), должны были соединиться вместе и образовать тот газ (в основном водород), из которого сформировались звезды спустя примерно 10^8 лет после Большого взрыва.

Звезды, однако, формировались постепенно. После дальнейшего расширения и охлаждения газа, его концентрация в некоторых частях вселенной могла немного увеличиться, что было необходимо для того, чтобы гравитационное притяжение в этой области начало доминировать над всеобщим расширением. Здесь мы подходим к еще нерешенному и спорному вопросу о действительном механизме формирования галактик, и о характере тех начальных неоднородностей, которые обеспечивают возможность формирования галактик. Я не собираюсь обсуждать сейчас эти вопросы. Мы только примем как факт, что в начальном распределении газа должны были иметь место некоторые неоднородности, и в определенный момент вступил в действие определенный механизм гравитационной конденсации, который обеспечил формирование галактик, со всеми сотнями и тысячами миллионов составляющих их звезд!

Мы установили, откуда взялся рассеянный газ. Он возник из того самого протошара, которым, собственно, и являлся Большой взрыв. Второе начало термодинамики в той детальной форме, в которой оно дошло до нас, обязано своим существованием факту удивительно равномерного распределения этого газа в пространстве после того, как стали возможными процессы гравитационной конденсации, повышающие полную энтропию. Насколько же однородно распределено вещество в настоящей вселенной? Мы уже обращали внимание на то, что звезды во вселенной собраны в галактики. Галактики, в свою очередь, группируются в скопления галактик, а скопления — в так называемые сверхскопления. Есть даже

некоторые свидетельства в пользу того, что эти сверхскопления собираются в огромные группировки, известные как комплексы сверхскоплений. Необходимо заметить, однако, что все эти неоднородности и скопления — «совершенные пустяки» в сравнении с поразительной однородностью структуры вселенной в целом. При этом, чем глубже в прошлое мы бы могли проникнуть и чем больший объем вселенной мы бы могли исследовать, тем все более однородной оказывалась бы вселенная. Чернотельное фоновое излучение дает самое убедительное свидетельство в пользу такого вывода. Отсюда следует, в частности, что когда возраст вселенной равнялся всего одному миллиону лет, на тех масштабах, которые сейчас расширились до 10^{23} километров (такое расстояние от нас включает около 10^{10} галактик) вселенная и все составляющее ее вещество были однородны с точностью до одной стотысячной (см. Дэвис и др. [1987]). Вселенная, несмотря на взрывообразный характер своего рождения, была в высокой степени однородной на своих самых ранних стадиях.

Итак, мы поняли, что это был горячий протошар, который так однородно распределил газ по всему пространству. Именно к этому моменту нас привели наши исследования.

Объясняется ли второе начало Большим взрывом?

Но закончились ли на этом наши поиски оснований для второго начала термодинамики?

Можно ли объяснить тот интригующий факт, что начальная энтропия вселенной была чрезвычайно мала — а именно он и дает нам второе начало — тем обстоятельством, что вселенная началась с Большого взрыва? После некоторых размышлений мы обнаружим, что такое объяснение содержит в себе парадокс. И потому оно никак не может быть окончательным ответом. Вспомним, что первичный протошар представлял собой некоторое *тепловое* состояние, а именно, горячий расширяющийся газ в тепловом равновесии. Вспомним также, что термин «тепловое равновесие» относится к состоянию с максимальной энтропией. (Именно так мы определяли максимум энтропии для газа в ящике.) Но второе начало требует, чтобы в начальном состоянии наша вселенная, напротив, находилась, в своего рода минимуме, но не в максимуме!

Где же ошибка? Один из «стандартных» ответов примерно таков:

> Действительно, в самом начале протошар пребывал в тепловом равновесии, но вселенная в то время была совсем маленькой. Этот протошар находился в состоянии с максимумом энтропии, разрешенной для вселенной с крошечными размерами, но такая энтропия оказалась бы совершенно ничтожной по сравнению с энтропией, разрешенной для вселенной с ее сегодняшними размерами. По мере расширения вселенной, соответствующий максимум энтропии также возрастал с увеличением размеров вселенной, но реальное значение энтропии вселенной сильно отставало от этого максимума. Следовательно, второе начало появляется благодаря тому, что реальная энтропия всегда пытается догнать свой разрешенный максимум.

Однако, небольшой анализ показывает, что такое объяснение не может быть правильным. Если бы оно было верным, то в случае (пространственно замкнутой) вселенной, которая испытывает в конце большой коллапс, это объяснение можно было бы применить снова в обратном направлении времени. Когда вселенная снова достигнет крошечных размеров, для максимума ее энтропии снова будет очень низкий потолок. То же самое рассуждение, благодаря которому мы получили низкую энтропию на ранних этапах расширяющейся вселенной, должно быть применено на конечных этапах существования сжимающейся вселенной. Именно некоторый предел на низкую энтропию в «начале времен» был тем, что дало нам второе начало, согласно которому энтропия вселенной возрастает со временем. Но если тот же самый предел должен быть применен и в «конце времен», то мы бы обнаружили вопиющее несоответствие со вторым началом.

Конечно, не исключено, что наша действительная вселенная никогда не будет сжиматься подобным образом. Возможно,

что мы живем во вселенной с нулевой средней пространственной кривизной (Евклидов случай), или с отрицательной кривизной (случай Лобачевского). Или может оказаться, что мы живем во вселенной (с положительной кривизной), которая обречена на схлопывание, но оно будет иметь место в столь отдаленном будущем, что никакие отклонения от второго начала в настоящий момент нам еще не видны, несмотря на то, что, вообще говоря, вся энтропия вселенной должна будет в некоторый момент времени начать уменьшаться до чрезвычайно малых величин, и второе начало — в том смысле, в котором мы понимаем его сегодня — будет сильно нарушено.

На самом деле, имеются веские основания сомневаться в подобной смене энтропии в сжимающейся вселенной. Самые убедительные из них связаны с загадочными объектами, именуемыми черными дырами. На примере черной дыры мы имеем микрокосмос сжимающейся вселенной, поэтому, если бы смена энтропии в такой вселенной действительно имела бы место, то в окрестности черной дыры должно было бы наблюдаться нарушение второго начала. Есть, однако, все основания полагать, что второе начало имеет прочную власть и над черными дырами. В процессе нашего обсуждения энтропии нам не удастся избежать вопросов, связанных с теорией черных дыр, поэтому нам просто необходимо познакомиться с этими странными объектами поближе.

Черные дыры

Начнем с теоретических предсказаний дальнейшей судьбы нашего Солнца. К настоящему моменту оно уже просуществовало около пяти миллиардов лет. В ближайшие 5–6 миллиардов лет оно начнет увеличиваться в размерах и будет непрерывно раздуваться до тех пор, пока его поверхность не достигнет где-то орбиты Земли. Тогда оно превратится в звезду, называемую *красным гигантом*. На небосводе можно обнаружить множество красных гигантов, из которых наиболее известны Альдебаран в созвездии Тельца и Бетельгейзе в созвездии Ориона. Пока поверхность красного гиганта расширяется, в самой его сердцевине

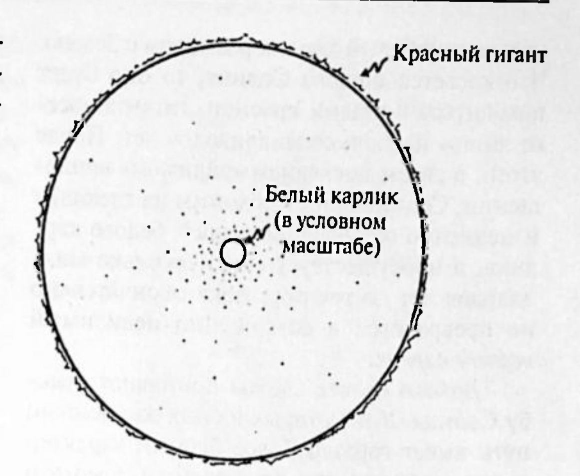


Рис. 7.12. Красный гигант с белым карликом в своей сердцевине

находится чрезвычайно плотная концентрация материи, которая непрерывно увеличивается в размерах. Это плотное ядрышко имеет ту же природу, что и звезда, называемая *белым карликом* (рис. 7.12).

Сами по себе белые карлики — это самые настоящие звезды, вещество которых, правда, спрессовано до такой степени, что теннисный шарик, заполненный им, весил бы несколько сотен тонн! Их число на небосводе довольно велико: примерно десять процентов всех светящихся звезд Млечного Пути приходится на белые карлики. Самый знаменитый из них — спутник Сириуса, чья невообразимо высокая плотность представляла большую загадку для наблюдательной астрономии в начале XX века. Однако позже именно эта звезда превосходным образом подтвердила справедливость физической теории (выдвинутой Р. Х. Фаулером примерно в 1926 году), согласно которой некоторые звезды и в самом деле могут обладать колоссальной плотностью, при которой они удерживаются в равновесии «давлением электронного вырождения». Это означает, что от гравитационного коллапса такую звезду спасает только квантовомеханический принцип запрета Паули (с. 246), примененный к электронам.

Любой красный гигант имеет ядро в виде белого карлика, и это ядро постоянно затягивает в себя вещество из основного тела звезды. В конце концов, красный гигант будет целиком поглощен своим ядром-паразитом и в результате останется самый

настоящий белый карлик размером с Землю. Что касается нашего Солнца, то оно будет находиться в стадии красного гиганта «всего лишь» несколько миллиардов лет. После этого, в своем последнем «видимом» воплощении, Солнце станет похожим на тлеющие и медленно остывающие угли[6] белого карлика, и просуществует еще несколько миллиардов лет до тех пор, пока окончательно не превратится в совершенно невидимый *черный карлик*.

Далеко не все звезды повторяют судьбу Солнца. У некоторых из них жизненный путь имеет гораздо более бурный характер и определяется так называемым *пределом Чандрасекара*: максимально возможной массой белого карлика. Согласно вычислениям, проведенным еще в 1929 году Субраманьяном Чандрасекаром, белые карлики не могут существовать, если их масса превышает шесть пятых массы Солнца. (В то время он был еще совсем молодым индийским студентом, и свои вычисления делал во время морского путешествия из Индии в Англию.) Затем, где-то в 1930 году, эти вычисления были проделаны (независимо) еще раз советским теоретиком Львом Давидовичем Ландау. Современное уточненное значение предела Чандрасекара составляет примерно

$$1{,}4M_\odot,$$

где $M_\odot$ — масса Солнца.

Заметим, что предел Чандрасекара совсем ненамного превосходит солнечную массу, и в то же время известно множество звезд, обладающих значительно бо́льшими массами. А как сложится судьба звезды, обладающей массой, скажем, $2M_\odot$? В соответствии с принятой теорией, такая звезда будет так же раздуваться и превратится в красный гигант, а ее ядро — белый карлик — будет постепенно набирать свою массу, — т. е. в точности так, как было описано выше. Однако, в некоторый критический момент, это ядро достигнет предела Чандрасекара, и принцип запрета Паули уже не сможет обеспечить давление, необходимое для компенсации чудовищных сил гравитации [8]. Примерно в этот момент ядро начнет катастрофически коллапсировать внутрь, что приведет, в свою очередь, к значительному повышению давления и температуры. Начнутся интенсивные ядерные реакции, и колоссальное количество энергии выделится из ядра в виде нейтрино. Они разогреют внешнюю оболочку коллапсирующей звезды, а затем последует грандиозный взрыв. Звезда превратится в сверхновую!

А что произойдет далее со все еще коллапсирующим ядром? Теория утверждает, что оно достигнет таких немыслимых плотностей, которые намного превосходят даже плотность белого карлика. Тогда коллапс ядра остановится и оно станет *нейтронной звездой* (с. 282), в которой теперь уже *давление нейтронного вырождения* (т. е. принцип Паули, примененный к нейтронам) будет удерживать ее в равновесии. Ее плотность такова, что теннисный мячик, сделанный из вещества нейтронной звезды, весил бы как астероид Гермес (или как марсианская луна Демос). Такую плотность имеет само ядерное вещество! (По сути дела, можно сказать, что нейтронная звезда представляет собой гигантское атомное ядро, радиусом с десяток километров, что, впрочем, совсем немного по звездным меркам!) Но здесь вступает в силу *новый* предел, аналогичный пределу Чандрасекара (и называемый *пределом Ландау—Оппенгеймера—Волкова*), который приближенно (по уточненным современным данным) составляет

$$2{,}5M_\odot,$$

и по превышении которого равновесие нейтронной звезды невозможно.

А что случится с коллапсирующим ядром, если масса исходной звезды будет настолько велика, что даже и *этот* предел будет превышен? Кстати говоря, известно много звезд, масса которых заключена в пределах от $10M_\odot$ до $100M_\odot$. Маловероятно, что все они в процессе взрыва сверхновой сбрасывают столь большую массу, что ядро-остаток неизменно оказывается с массой ниже верхнего предела для нейтронной звезды. Вместо этого мы, скорее всего, получим *черную дыру*.

---

[6] В действительности, на этой конечной стадии карлик будет светиться как красная звезда, но то, что называют «красными карликами», относится к звездам совсем другого типа.

Что же такое черная дыра? Это — область пространства (или пространства-времени), в пределах которой гравитационное поле настолько сильно, что даже свет не способен вырваться из нее. Вспомним, что в силу принципа относительности скорость света является предельной: ни один материальный объект или сигнал не может превысить ее локальное значение (с. 183). Следовательно, если даже свет не может вырваться из черной дыры, то из нее не сможет выбраться наружу вообще *ничего*.

Читатель, возможно, знаком с понятием *второй космической скорости*. Это та минимальная начальная скорость, которую должен иметь объект, чтобы он мог удалиться от некоторого массивного тела на сколь угодно большое расстояние. Пусть массивное тело — это Земля, тогда вторая космическая скорость для нее будет составлять примерно 40 000 км/ч километров в час. Камень, брошенный с земной поверхности (в любом направлении) со скоростью, превышающей это значение, навсегда покинет Землю (мы, конечно, пренебрегаем силами сопротивления земной атмосферы). Если же камень бросить со скоростью, меньшей этого значения, он упадет обратно на Землю. (Поэтому *неверно* утверждение, что «все брошенное вверх, обязательно упадет вниз»; это будет справедливым только в том случае, когда скорость бросания меньше второй космической!) Для Юпитера вторая космическая скорость равна 220 000 км/ч; а для Солнца она будет составлять уже 2 200 000 км/ч. Теперь представим себе, что вся солнечная масса оказалась сосредоточенной в сфере радиуса в *одну четверть* от его истинного значения. В этом случае вторая космическая скорость увеличится в *два раза* по сравнению с исходным значением. А если бы вся масса Солнца оказалась сосредоточенной в еще меньшем объеме, скажем, внутри сферы радиуса в *одну сотую* от его истинного значения, вторая космическая скорость увеличилась бы в *десять раз*. Мы можем представить себе, таким образом, достаточно массивное тело малых размеров, для которого вторая космическая скорость превышает даже скорость света! Когда это происходит в действительности, мы и имеем черную дыру [9].

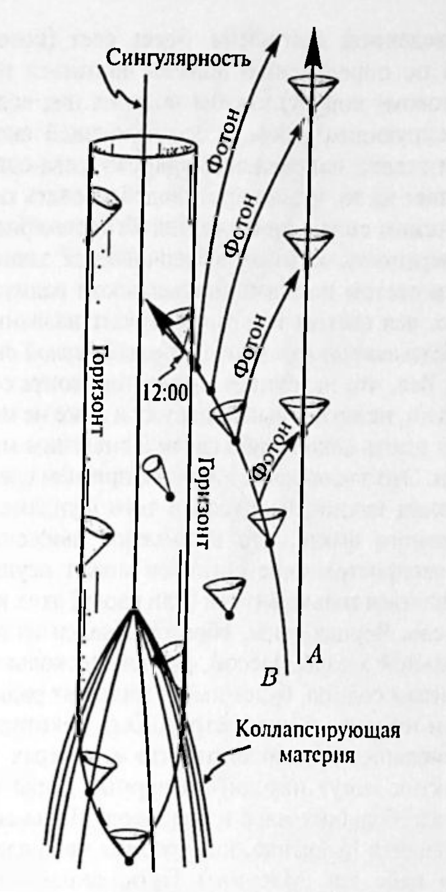


Рис. 7.13. Пространственно-временная диаграмма, демонстрирующая коллапс в черную дыру. Шварцшильдовский радиус обозначен как «горизонт»

На рис. 7.13 я изобразил пространственно-временную диаграмму, показывающую коллапс тела, который приводит к образованию черной дыры (для простоты я предположил, что в процессе коллапса сохраняется сферическая симметрия тела и убрал одно пространственное измерение). На рисунке изображены также световые конусы, которые, как мы помним из обсуждения общей теории относительности в главе 5 (см. с. 193), абсолютным образом ограничивают допустимые движения тел и распространение сигналов. Заметим, что эти конусы начинают немного наклоняться внутрь и чем ближе к центру, тем этот наклон становится более и более значительным.

Существует некоторое критическое расстояние от центра, называемое *шварцшильдовским радиусом*, на котором внешние границы конусов становятся *вертикальными* на

приведенной диаграмме. Здесь свет (который по определению должен двигаться по световому конусу) как бы зависает над коллапсирующим телом, а составляющей скорости света, направленной наружу, едва-едва хватает на то, чтобы противодействовать гигантским силам притяжения. Та трехмерная поверхность, которая вычерчивается зависшим светом на шварцшильдовском радиусе (т. е. вся световая история) носит название (*абсолютного*) *горизонта событий черной дыры*. Все, что находится внутри горизонта событий, не может выйти наружу и даже не может иметь какой-либо связи с внешним миром. Это заключение является прямым следствием наклона конусов и того фундаментального факта, что возможное движения и распространение сигналов может осуществляться только внутри (или вдоль) этих конусов. Черная дыра, образовавшаяся из начальной звезды массой, равной нескольким массам солнца, будет иметь горизонт радиусом несколько километров. Есть некоторые основания предполагать, что в центрах галактик могут находиться черные дыры гораздо больших масс и размеров. Наша собственная Галактика, которую мы наблюдаем на небе как Млечный Путь, вполне может содержать черную дыру, имеющую массу около миллиона солнечных масс и, следовательно, радиус горизонта — несколько миллионов километров.

Реальное материальное тело, которое коллапсирует с образованием черной дыры, в конце концов, целиком окажется внутри своего горизонта и, следовательно, потеряет всякую связь с внешним миром. Далее мы в общих чертах проследим вероятную судьбу этого тела, а в данный момент для нас будет представлять интерес только геометрия пространства-времени, порожденная этим коллапсом, которая приводит к весьма любопытным следствиям.

Вообразим, что некий отважный (а, может быть, безрассудный?) астронавт B собрался совершить путешествие в нутро большой черной дыры, а его менее смелый (а, может быть, просто более осторожный?) коллега A остается при этом за пределами горизонта событий. Предположим, что A намеревается держать B в поле своего зрения до тех пор, пока это в принципе возможно. Что же увидит A? Глядя на рис. 7.13, нетрудно сообразить, что ту часть истории B (т. е. мировой линии B), которая лежит *внутри* горизонта, A не увидит никогда, в то время как часть, лежащую *снаружи* горизонта, A рано или поздно увидит целиком, хотя те точки истории B, которые непосредственно предшествуют моменту его прохождения через горизонт, будут наблюдаться A спустя все большее и большее время ожидания. Пусть B проходит горизонт в тот момент, когда его собственные часы показывают 12 часов. Само это событие A не зафиксирует никогда, но события, соответствующие показаниям часов 11 : 30, 11 : 45, 11 : 52, 11 : 56, 11 : 58, 11 : 59, 11 : $59\frac{1}{2}$, 11 : $59\frac{3}{4}$, 11 : $59\frac{7}{8}$ и т. д. A будет наблюдать совершенно определенно (причем через приблизительно равные интервалы времени по показаниям часов A). В принципе, B будет все время находиться в поле зрения A как бы навечно зависшим над горизонтом черной дыры, а часы B будут отсчитывать время все медленнее и медленнее по мере приближения к роковой отметке 12 : 00, но никогда не покажут этого значения. Но, в действительности, образ B, который видит A, очень быстро станет неясным и трудноразличимым. Это происходит потому, что образ B, который будет наблюдать A в течение всего оставшегося времени наблюдения, будет формироваться лишь светом, испущенным из того крошечного участка мировой линии B, который непосредственно примыкает извне к горизонту. В результате B просто исчезнет из поля зрения A, и то же самое окажется верным и для исходного коллапсирующего тела. Все, что увидит A, будет выглядеть, в конце концов, действительно как какая-то «черная дыра»!

А что можно сказать о несчастном B? Каковы будут *его* ощущения? Необходимо сразу же подчеркнуть, что в момент пересечения горизонта с B ничего особенного не произойдет. Он смотрит на свои часы около 12 и видит, что минута за минутой следуют обычным порядком: 11 : 57, 11 : 58, 11 : 59, 12 : 00, 12 : 01, 12 : 02, 12 : 03, … . Промежуток времени около 12 : 00 не содержит ничего необычного. B может обернуться, посмотреть на A и убедиться, что A все время остается в поле его

зрения. *B* может также посмотреть на часы *A* и также убедиться, что для него они идут обычным образом. Таким образом, *B* никак не может узнать о своем пересечении горизонта, если только не проделает для этого специальных *расчетов* [10]. Горизонт оказался предельно коварным! После пересечения горизонта, у *B* уже не остается никаких шансов выйти наружу. Он обнаружит, что окружающая его часть вселенной сжимается, и что довольно скоро ему предстоит испытать свой собственный «большой коллапс»!

А может быть, и не только его собственное. Все вещество того первоначального тела, из которого образовалась черная дыра, будет, в некотором смысле, разделять судьбу *B* и испытывать такой же коллапс. Более того: в случае, если вселенная *снаружи* дыры пространственно замкнута, так что вся внешняя материя тоже оказывается вовлеченной в глобальный большой коллапс, то этот коллапс должен оказаться, в определенном смысле, «тем же самым», что и «собственный» коллапс *B* [7]).

Но несмотря на столь безрадостный конец *B*, физика, с которой он будет иметь дело вплоть до гибельной точки, будет той самой, которую мы с вами хорошо знаем и понимаем. В частности, нет никаких оснований предполагать нарушение второго начала термодинамики, тем более предполагать, что полностью обратится монотонный рост энтропии. Второе начало будет действовать внутри черной дыры точно также, как и везде. Энтропия в окрестности *B* будет продолжать возрастать, вплоть до самого момента его окончательного коллапса.

Чтобы разобраться, каким образом энтропия «большого коллапса» («собственного» или «всеохватывающего») может быть чрезвычайно высокой, в то время как энтропия большого взрыва может оказаться при этом намного меньше, нам следует немного

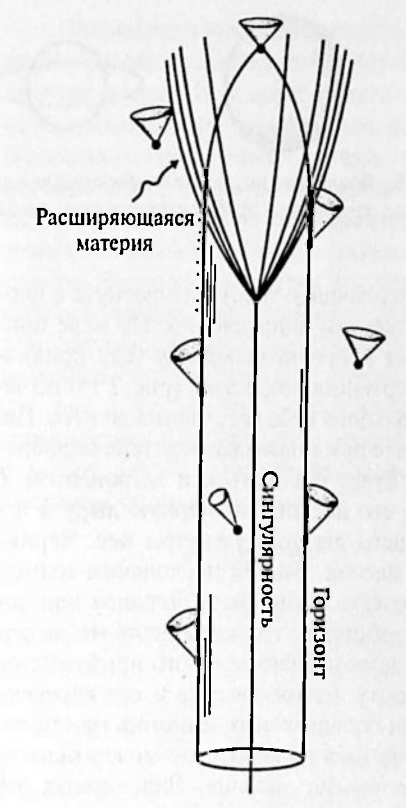


Рис. 7.14. Гипотетическая пространственно-временная конфигурация: белая дыра, эволюция которой приводит к расширяющейся материи (эта ситуация является обращением во времени рис. 7.13)

глубже вникнуть в свойства геометрии пространства-времени черной дыры. Перед тем, как мы этим займемся, читатель должен взглянуть на рис. 7.14, на котором показано гипотетическое временнóе обращение черной дыры, называемое *белой дырой*. (Скорее всего, белых дыр в природе не существует, но их теоретическая возможность будет иметь для нас большое значение в дальнейшем.)

Структура пространственно-временных сингулярностей

Вспомним из главы 5 (с. 191), как кривизна пространства-времени проявляется в *приливных эффектах*. Сферическая поверхность, образованная свободно падающими в гравитационном поле частицами некоторого большого тела, будет вытянута в одном направлении (вдоль линии, направленной

[7]) Делая подобное утверждение, я неявно ввожу следующие два допущения. Первое заключается в том, что возможному полному окончательному исчезновению черной дыры — с учетом ее (чрезвычайно медленного) хокинговского радиационного «испарения», которое мы рассмотрим чуть позже (см. с. 298) — будет предшествовать окончательный коллапс вселенной; второе допущение — (весьма правдоподобное), известно под названием «космическая цензура» (с. 198).

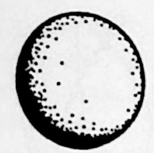
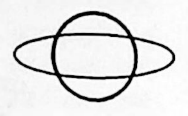
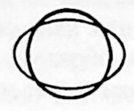
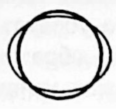
   

Рис. 7.15. Приливное воздействие, оказываемое сферическим притягивающим телом, возрастает по мере того, как другое тело приближается к нему, по закону обратного куба расстояния между центрами тел

на притягивающее тело) и сплюснута в перпендикулярном направлении. По мере приближения к притягивающему телу приливная деформация возрастает (рис. 7.15) по закону обратного куба расстояния до него. Нарастающее приливное воздействие подобного рода будет ощущаться и астронавтом *B* по мере его падения на черную дыру и последующего движения внутри нее. Черные дыры с массой, равной нескольким солнечным, оказывали бы столь большое приливное воздействие, что космонавт не выдержал бы даже незначительного приближения к горизонту, не говоря уже о его пересечении. Для больших дыр величина приливного воздействия на горизонте может оказаться существенно меньше. Для черных дыр с массой в миллион солнечных, одна из которых, как предполагают астрономы, находится в центре нашей Галактики — Млечного Пути, — приливное воздействие на горизонте, испытываемое астронавтом, было бы ничтожно малым, так что он, в худшем случае, ощутил бы лишь небольшой дискомфорт. Однако, это приливное воздействие менялось бы по мере дальнейшего падения астронавта внутри дыры, так что за какие-то секунды оно достигло бы, в конце концов, бесконечной величины! И не только тело бедного астронавта оказалось бы разорванным на кусочки этой очень быстро возрастающей приливной силой, но, как в ускоренном кино, оказались бы разорванными и молекулы, из которых это тело состоит, потом составляющие эти молекулы атомы, их ядра, и, в конце концов, вообще все какие только есть субатомные частицы! Таким образом, «коллапс» разрушает все до основания.

При этом разрушается не только материя, но даже и само пространство-время прекращает свое существование. Такая окончательная катастрофа называется *пространственно-временной сингулярностью*. Читатель, конечно, может задаться справедливым вопросом, откуда мы знаем, что подобные катастрофы должны иметь место, и при каких обстоятельствах материю и пространство-время ожидает такая судьба. Вывод о неизбежности пространственно-временной сингулярности следует из классических уравнений общей теории относительности и оказывается справедливым при любых условиях, в которых находится уже сформировавшаяся черная дыра. Первоначальная модель Оппенгеймера и Снайдера (Оппенгеймер, Снайдер [1939]) как раз и демонстрировала поведение подобного типа. Долгое время, однако, астрофизики питали надежду, что такое сингулярное поведение является артефактом специальной симметрии, которая допускалась в этой модели с самого начала. Предполагалось, что в реалистичном (асимметричном) случае коллапсирующая материя могла бы скручиваться каким-то другим способом, а затем снова вырываться наружу. Но эти надежды исчезли после того, как было проведено математическое исследование более общего характера, которое послужило основой для формулировки так называемых *теорем о сингулярности* (см. Пенроуз [1965]; Хокинг, Пенроуз [1970]). Эти теоремы утверждали, что в рамках классической общей теории относительности с разумными источниками гравитации, пространственно-временные сингулярности *неизбежны* в случае гравитационного коллапса.

Таким же образом, меняя направление времени, мы приходим к выводу о неизбежности соответствующей *начальной* пространственно-временной сингулярности, которую мы теперь представляем как Большой взрыв, в любой (надлежащим образом) расширяющейся вселенной. Только теперь, вместо окончательного *разрушения* пространства-времени и материи, эта сингулярность представляет собой *рождение* пространства-времени и материи. Может показаться, что име-

ется полная временная симметрия между этими двумя типами сингулярностей: *начальным* типом, при котором пространство-время и материя рождаются, и *конечным* типом, когда пространство-время и материя уничтожаются. Конечно, между этими двумя ситуациями действительно имеется важная аналогия, но исследуя их более детально, мы обнаружим, что они не являются *точными* копиями, обращенными во времени относительно друг друга. И для нас важно разобраться в тех различиях геометрического характера, которые имеются между ними, поскольку именно они оказываются ключевыми в понимании источника второго начала термодинамики!

Обратимся к наблюдениям нашего астронавта B, который отважился на самопожертвование ради науки. Он наблюдает приливные силы, которые очень быстро возрастают до бесконечности. Поскольку он путешествует в пустом пространстве, то он ощущает *деформирующие* эффекты, которые оставляют величины объемов неизменными и которые создаются частью тензора пространственно-временной кривизны, обозначенной мною как ВЕЙЛЬ (см. главу 5, с. 191). Другая часть тензора пространственно-временной кривизны, отвечающая за общее изменение объемов и называемая РИЧЧИ, обращается в нуль в пустом пространстве. Может оказаться, что B все же встретится с какой-нибудь материей в некоторый момент, но даже если это действительно произойдет (ведь, в конце концов, и сам астронавт состоит из материальных частиц), мы, вообще говоря, все равно обнаружим, что величина ВЕЙЛЬ будет намного *превосходить* величину РИЧЧИ. Таким образом, значение кривизны вблизи *конечной* сингулярности полностью определяется поведением тензора ВЕЙЛЬ. Этот тензор, вообще говоря, стремится к *бесконечности*:

$$\text{ВЕЙЛЬ} \to \infty$$

(хотя это стремление может иметь осциллирующий характер). Эта ситуация оказывается *типичной* для пространственно-временной сингулярности [11]. Такое поведение связано с *высокоэнтропийной* сингулярностью.

Однако в случае Большого взрыва, ситуация оказывается совершенно другой. Стандартная модель Большого взрыва выводится из рассмотренных нами ранее вселенных Фридмана—Робертсона—Уокера, обладающих высокой степенью симметрии. Здесь деформирующее приливное воздействие, связанное с тензором ВЕЙЛЬ, вообще *отсутствует*. Вместо него теперь имеется направленное внутрь симметричное ускорение, действующее на любую сферическую поверхность, состоящую из пробных частиц (см. рис. 5.26). Но это — результат воздействия тензора РИЧЧИ, а не тензора ВЕЙЛЬ. В любой ФРУ-модели всегда имеет место тензорное уравнение:

$$\text{ВЕЙЛЬ} = 0.$$

По мере того, как мы приближаемся к начальной сингулярности все ближе и ближе, мы обнаруживаем, что именно РИЧЧИ, а не ВЕЙЛЬ, становится бесконечным и, таким образом, именно РИЧЧИ, а не ВЕЙЛЬ, определяет начальную сингулярность. Значит, мы имеем дело с *низкоэнтропийной* сингулярностью.

Если мы исследуем сингулярность схлопывания в *точной* коллапсирующей ФРУ-модели, мы и здесь обнаружим, что в момент схлопывания ВЕЙЛЬ $= 0$, тогда как

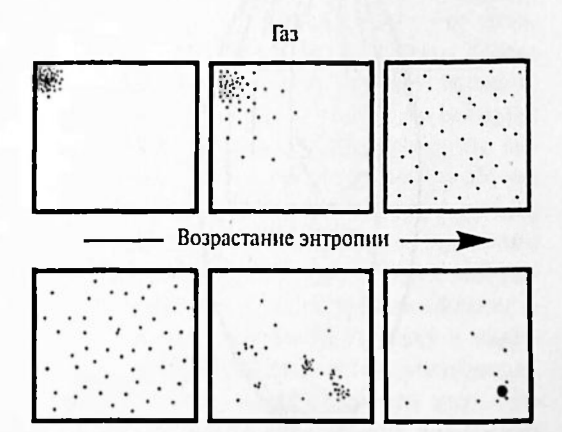


Рис. 7.16. Для обычного газа повышение энтропии связано с увеличением степени однородности его распределения внутри ящика. Для гравитирующих систем имеет место обратная ситуация. Высокая энтропия соответствует гравитационному конденсату, а максимальная — образованию черной дыры

РИЧЧИ стремится к бесконечности. Однако, эта особая ситуация дает нам совсем *не то*, что мы ожидаем от более реалистичной модели, в которой учитывается также и гравитационная конденсация. С течением времени вещество, находящееся первоначально в виде рассеянного газа, будет конденсироваться в звездные галактики. В этом процессе большое число звезд испытают гравитационное сжатие и превратятся в белые карлики, нейтронные звезды и черные дыры, а также в гигантские черные дыры, которые вполне могут образоваться в центрах галактик. Такого рода конденсация — особенно в случае черных дыр — связана с огромным возрастанием энтропии (рис. 7.16). Может показаться странным, на первый взгляд, что конденсированные состояния дают *большую* энтропию, чем состояния с однородным распределением, особенно если вспомнить, что для газа в ящике его конденсированные состояния (например, случай, когда весь газ собирается в одном из углов ящика) имели *низкую* энтропию, в то время как *однородное* распределение, соответствующее тепловому равновесию — имело *высокую* энтропию. При учете гравитации ситуация меняется на *обратную* благодаря универсальности гравитационного притяжения. С течением времени, конденсация становится все более и более сильной и, в конце концов, множество сконденсировавшихся черных дыр соединяет свои сингулярности в финальной сингулярности большого коллапса. Такая конечная сингулярность не имеет ничего общего с тем идеализированным большим коллапсом, который имеет место в коллапсирующей ФРУ-модели, где действовало ограничение ВЕЙЛЬ $= 0$. По мере накопления числа сконденсировавшихся объектов, тензор ВЕЙЛЬ имеет тенденцию непрерывно увеличиваться [12] и, вообще говоря, ВЕЙЛЬ $\to \infty$ в конечной сингулярности. Посмотрите на рис. 7.17, где показана полная история замкнутой вселенной в соответствии с этой общей картиной.

Мы видим теперь, как становится возможной ситуация, когда сжимающаяся вселенная может не обладать низкой энтропией. Та «малость» энтропии Большого взрыва, которая обеспечивает нам выполнение второго начала, не была, таким образом, следствием *одной только* «малости» вселенной в момент взрыва! Если бы мы обратили



Рис. 7.17. Полная история замкнутой вселенной, которая начинается с однородного низкоэнтропийного большого взрыва с ограничением ВЕЙЛЬ $= 0$ и заканчивается высокоэнтропийным большим коллапсом — представляющим собой сгущение большого числа черных дыр — с условием ВЕЙЛЬ $\to \infty$

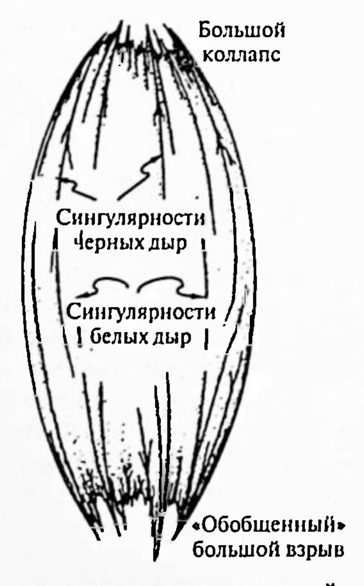


Рис. 7.18. Если убрать ограничение ВЕЙЛЬ $= 0$, то большой взрыв получится тоже высокоэнтропийным, с условием ВЕЙЛЬ $\to \infty$. Такая вселенная была бы сплошь испещрена белыми дырами и в ней не выполнялось бы второе начало термодинамики — в полном противоречии с нашим опытом

во времени картину большого коллапса, к которой только что пришли, мы бы получили «большой взрыв» с чрезвычайно *высокой* энтропией, где не было бы второго начала! По некоторым причинам, вселенная возникла в особом (низкоэнтропийном) состоянии, на которое было наложено условие типа ВЕЙЛЬ = 0 для ФРУ-моделей. И если бы подобного рода ограничение не имело места, то «намного более вероятной» могла бы оказаться ситуация, в которой *как* начальная, *так* и конечная сингулярности были бы высокоэнтропийного типа ВЕЙЛЬ → ∞ (рис. 7.18). В такой гипотетической вселенной, конечно же, не нашлось бы места для второго начала термодинамики!

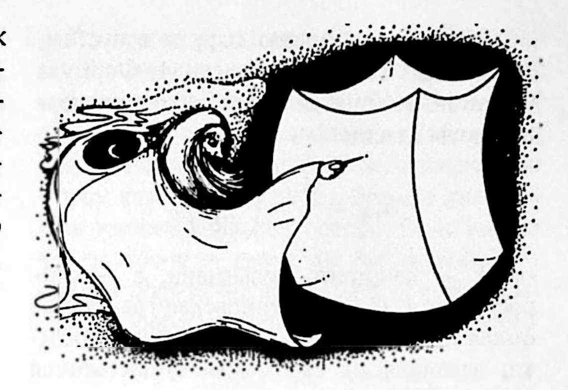


Рис. 7.19. Для сотворения вселенной, близкой по своим свойствам к той, в которой мы живем, Творец ограничивает свой выбор исчезающе малым объемом в фазовом пространстве возможных вселенных, в рассматриваемом случае — всего около $1/10^{10^{123}}$ объема всего пространства. (Этот объем и нацеленная на него булавка показаны без соблюдения масштабов!)

Насколько особым был Большой взрыв?

Попробуем разобраться с вопросом о том, насколько ограничивающим для Большого взрыва было условие типа ВЕЙЛЬ = 0. Для простоты (как и ранее) мы будем считать вселенную замкнутой. Для того чтобы составить ясную и конкретную картину, далее мы везде будем полагать, что число *барионов* B — т. е. общее число протонов и нейтронов, во вселенной составляет примерно

$$B = 10^{80}.$$

(Не существует каких-то особых оснований для выбора именно этого значения, кроме тех эмпирических данных, которые приводят к нему как к *нижней* оценке B. Эддингтон однажды заявил, что вычислил B *точно* и полученное им значение оказалось близким к приведенному выше! Кажется, что сейчас уже никто не принимает всерьез эти вычисления, но значение 10^{80} надежно утвердилось.) Если бы мы взяли *большее* значение B (в действительности может оказаться, что $B = \infty$), то величины, полученные нами в этом случае, оказалась бы еще *поразительнее* тех (и без того весьма экстраординарных чисел), к которым мы через несколько шагов придем!

Попробуем представить себе фазовое пространство (с. 169) *всей* вселенной! Каждая точка этого пространства потенциально соответствует определенному начальному состоянию, из которого вселенная могла начинать свою эволюцию. На рис. 7.19 мы условно изображаем Творца, который в своей деснице держит «булавку», чтобы отметить ею некую точку нашего фазового пространства. Каждое положение булавки соответствует творению особой вселенной. Точность, с которой Творец создает какую-либо вселенную, напрямую связана с энтропией этой вселенной. Создать вселенную с высокой энтропией было бы относительно «легко», поскольку в этом случае в распоряжении Творца имеется большой объем фазового пространства, в который надо указать булавкой. (Напомним, что энтропия пропорциональна логарифму объема соответствующего фазового пространства.) Но чтобы создать вселенную в состоянии с низкой энтропией — так, чтобы в ней выполнялось второе начало термодинамики, — Творец должен направить булавку в гораздо меньший объем фазового пространства. Насколько малым должен быть этот объем, чтобы в результате творения получилась вселенная, напоминающая по своим свойствам ту, в которой мы живем? Для ответа на этот вопрос, мы должны обратиться к замечательной формуле, выведенной Якобом Бекенштейном [1972] и Стивеном Хокингом [1975], которая говорит о том, чему должна быть равна энтропия *черной дыры*.

Рассмотрим черную дыру и допустим, что площадь ее горизонта есть A. Формула Хокинга—Бекенштейна для энтропии черной дыры гласит:

$$S_{\text{ч.д.}} = \frac{A}{4} \times \left(\frac{kc^3}{G\hbar}\right),$$

где k — константа Больцмана, c — скорость света, G — ньютоновская гравитационная постоянная и $\hbar$ — постоянная Планка, деленная на 2π. Самая существенная часть этой формулы заключена во множителе $A/4$. Часть, стоящая в скобках, содержит только необходимые для соблюдения размерности физические константы. Таким образом, энтропия черной дыры оказывается пропорциональной площади ее поверхности. Для сферически симметричной черной дыры эта площадь оказывается пропорциональной квадрату массы этой дыры:

$$A = m^2 \times 8\pi \frac{G^2}{c^4}.$$

Объединяя это с формулой Бекенштейна—Хокинга, мы получаем, что энтропия черной дыры пропорциональна квадрату ее массы:

$$S_{\text{ч.д.}} = m^2 \times 2\pi \frac{kG}{\hbar c}.$$

Таким образом, *энтропия, приходящаяся на единицу массы* ($S_{\text{ч.д.}}/m$) черной дыры, пропорциональна ее массе и оказывается тем больше, чем больше черная дыра. Следовательно, для заданной массы или, эквивалентно, — согласно формуле Эйнштейна $E = mc^2$, — для заданной *энергии*, наибольшая энтропия достигается тогда, когда вся материя сколлапсирует в черную дыру! Более того, энтропия системы двух черных дыр существенно возрастает, когда эти дыры сливаются в одну! Гигантские черные дыры, типа тех, которые, как полагают, находятся в центрах галактик, заключают в себе колоссальное количество энтропии — намного превосходящее те ее значения, которые встречаются в других физических ситуациях.

Утверждение о том, что максимум энтропии достигается при коллапсе всей массы в черную дыру, требует небольшого пояснения. Анализ термодинамики черных дыр, проведенный Хокингом, показывает, что с любой черной дырой можно связать некоторую ненулевую температуру. Одним из следствий этого является тот факт, что в состоянии с максимальной энтропией в черной дыре не может быть заключена вся масса-энергия; максимум энтропии достигается, когда черная дыра приходит в тепловое равновесие с «тепловым резервуаром излучения». Температура этого излучения оказывается действительно ничтожной для черных дыр с любым разумным размером. Так, к примеру, для черной дыры с массой порядка массы Солнца эта температура оказалась бы равной примерно 10^{-7} K, что значительно ниже температур, достигнутых в настоящее время в лабораториях, и намного меньше температуры 2,7 K межгалактического пространства. Для черных дыр больших размеров температура Хокинга оказывается еще меньшей!

Эта температура могла бы оказаться существенной для нашего обсуждения только в том случае, если либо (а) во вселенной существуют намного меньшие черные дыры, которые называют черными мини-дырами; либо (б) вселенная не успеет полностью сколлапсировать за время, меньшее *хокинговского времени испарения* — времени, за которое черная дыра полностью испаряется. Относительно (а) надо заметить, что черные мини-дыры могут возникнуть лишь в случае особенно хаотичного Большого взрыва. В нашей вселенной их не может быть очень много, в противном случае они бы уже както проявили бы себя; более того, согласно излагаемой мной здесь точки зрения, их вообще не должно быть.

Что же касается (б), то для черной дыры с солнечной массой хокинговское время испарения имело бы величину, превосходящую нынешний возраст вселенной где-то в 10^{54}; а для черных дыр бóльших размеров оно оказалось бы еще более продолжительным. Таким образом, вряд ли эффект испарения может существенно изменить наши предыдущие рассуждения.

Чтобы иметь некоторое представление о гигантских величинах энтропии черных дыр, рассмотрим чернотельное фоновое излучение с температурой 2,7 K, которое, как долго казалось, давало наибольший вклад в энтропию вселенной. Астрофизики были

просто ошарашены огромным количеством энтропии, заключенным в этом излучении, которое намного превосходило все значения энтропии, с которыми приходилось сталкиваться в других ситуациях (например, на Солнце). Энтропия фонового излучения составляет примерно 10^8 на один барион (здесь я снова перехожу к «естественной системе единиц», в которых постоянная Больцмана равна единице). (По сути, это означает, что на каждый барион приходится 10^8 фотонов фонового излучения.) Таким образом, если всего имеется 10^{80} барионов, то для полной энтропии фонового излучения во вселенной мы имели бы величину

$$10^{88}.$$

Несомненно, что если бы не было черных дыр, то эта величина представляла бы собой практически всю энтропию вселенной, поскольку энтропия фонового излучения намного превосходит энтропию всех других обычных процессов. Так, например, энтропия, приходящаяся на один барион на Солнце, оказывается порядка единицы. С другой стороны, по меркам черных дыр, энтропия фонового излучения — это просто «писк комара». Для черной дыры в одну солнечную массу формула Бекенштейна—Хокинга дает нам значение энтропии около 10^{20} на один барион (в естественных единицах). И даже если бы вселенная состояла всего-навсего из одной черной дыры с массой Солнца, полная энтропия оказалась бы уже намного превосходящей приведенное ранее значение, а именно, была бы равной

$$10^{100}.$$

Конечно, вселенная не устроена таким образом, но эта цифра определенно свидетельствует о том, насколько несущественной становится энтропия фонового излучения, когда мы начинаем учитывать влияние вездесущей гравитации.

А теперь попробуем сделать более реалистичную оценку. Вместо того, чтобы заселять наши галактики одними только черными дырами, примем, что эти галактики состоят, в основном, из обычных звезд — примерно 10^{11} штук в каждой и еще содержат в своей сердцевине около миллиона (10^6) черных дыр с массой солнца (что было бы вполне правдоподобно для нашей собственной Галактики — Млечного Пути). Вычисления показывают, что энтропия, приходящаяся на один барион, оказалась бы в этом случае существенно больше даже того огромного значения, которое было только что получено — она стала бы равной 10^{21}, что для полной энтропии дает (в естественных единицах) величину, равную примерно

$$10^{101}.$$

Мы можем предположить, что, по истечении достаточно большого промежутка времени, подавляющая часть галактических масс окажется захваченной черными дырами в центрах галактик. Когда это произойдет, энтропия в расчете на один барион станет равной 10^{31}, что дает чудовищное значение для полной энтропии:

$$10^{111}.$$

Мы, однако, рассматриваем замкнутую вселенную, которая, в конце концов, должна сколлапсировать; и было бы вполне разумно оценить энтропию конечного коллапса, используя формулу Бекенштейна—Хокинга и полагая при этом, что вся вселенная в момент коллапса представляет собой одну черную дыру. Такая оценка дает величину энтропии на один барион около 10^{43} и совершенно немыслимую величину полной энтропии для конечного коллапса:

$$10^{123}.$$

Это число мы будем рассматривать как некоторую оценку полного объема фазового пространства **V**, доступного для Творца, поскольку эта энтропия должна представлять собой логарифм объема (несомненно) наибольшей его части. Поскольку 10^{123} есть логарифм объема, сам объем должен представлять собой экспоненту от 10^{123}, т. е.

$$V = 10^{10^{123}}$$

в естественных единицах! (Некоторые особо внимательные читатели могли заметить, что я должен был написать $e^{10^{123}}$ — но для чисел такого порядка разница между основаниями e и 10 совершенно несущественна!) А каков был исходный объем фазового

пространства, на который должен был нацелиться Творец, чтобы сотворить вселенную, совместимую со вторым началом термодинамики? Оказывается, что совершенно не важно, какое выбрать значение

$$\mathbf{W} = 10^{10^{101}} \quad \text{или} \quad \mathbf{W} = 10^{10^{88}},$$

определяемое галактическими черными дырами или фоновым излучением соответственно, а, может быть, даже еще меньшее (и, на самом деле, более вероятное), которое могло иметь место в реальных условиях при Большом взрыве.

В любом случае, значение отношения $\mathbf{V}$ к $\mathbf{W}$ будет приблизительно:

$$\frac{\mathbf{V}}{\mathbf{W}} = 10^{10^{123}}.$$

(Проверьте сами: $10^{10^{123}} : 10^{10^{101}} = 10^{(10^{123}-10^{101})}$ дает с хорошим приближением $10^{10^{123}}$.)

Эта величина свидетельствует о том, насколько точным должен был быть замысел Творца: точность составляла примерно *одну* $10^{10^{123}}$*-ую*! Это поразительная точность. Полученную цифру нельзя даже полностью выписать в обычной десятичной системе исчисления: она представляла бы собой «1» с последующими 10^{123} нулями! Даже если бы мы были в состоянии записать «0» на каждом протоне и каждом нейтроне во вселенной, а также использовали бы для этой цели все остальные частицы, наше число, тем не менее, осталось бы недописанным.

Точность, необходимая для задания начальных условий вселенной, как видно, совершенно несоизмерима с той весьма высокой точностью, которая уже стала привычной, когда речь заходит о динамических уравнениях (Ньютона, Максвелла, Эйнштейна), управляющих поведением физических объектов в различных ситуациях.

Но почему же Большой взрыв был организован с такой высокой степенью точности, в то время как большой коллапс (или сингулярности черных дыр) должен быть совершенно хаотичным? Может показаться, что этот вопрос стоило бы переформулировать в терминах поведения ВЕЙЛЬ-части пространственно-временно́й кривизны в пространственно-временно́й сингулярности. Мы установили, что имеется ограничение

$$\text{ВЕЙЛЬ} = 0$$

(или нечто похожее) в сингулярностях начального типа, отсутствующее в конечных сингулярностях — и, кажется, именно оно отражает выбор Творцом соответствующей крошечной области фазового пространства. Предположение о том, что такое ограничение применимо к любой начальной (но не конечной!) сингулярности, я назвал бы Гипотезой Вейлевской Кривизны. Таким образом, напрашивается вывод, что нам осталось понять лишь одну вещь для окончательного разрешения вопроса о происхождении второго начала термодинамики, а именно: почему мы должны использовать такую несимметричную во времени гипотезу? [13]

Но как нам преодолеть это (последнее?) препятствие на пути к полному пониманию причины существования второго начала? Кажется, мы попали в безвыходное положение. Нам необходимо понять, почему пространственно-временны́е сингулярности имеют определенную структуру; но пространственно-временны́е сингулярности представляют собой как раз те области, в которых наше понимание физики достигает своих пределов. Этот тупик, связанный с существованием пространственно-временны́х сингулярностей, иногда сравнивают с другим тупиком: он имел место в начале XX века и был связан с проблемой устойчивости атомов (см. с. 209). В каждом из этих случаев хорошо обоснованная классическая теория приводит к ответу «бесконечность» и обнаруживает, тем самым, свою несостоятельность для решения соответствующей проблемы.

Сингулярный характер электромагнитного коллапса атомов был устранен *квантовой* теорией. Аналогично именно квантовая теория должна привести теперь к конечной теории, взамен «бесконечных» классических пространственно-временны́х сингулярностей, возникающих при гравитационном коллапсе звезд. Но это не может быть обычная квантовая теория. Это должна быть квантовая теория самой структуры пространства и времени. Такую теорию (в случае, если она вообще появится) следовало бы назвать «квантовой гравитацией». То, что ее пока нет, не является признаком недостатка усилий, опыта или изобретательности физиков: многие первоклассные

ученые с мировым именем пытались построить такую теорию, но безуспешно. Это тупик, к которому, в конце концов, пришли и мы в наших попытках понять направленность и течение времени.

Читатель может справедливо спросить: а что же, в таком случае, дало нам наше путешествие? В наших поисках понимания того, почему время кажется текущим только в одном направлении, нам пришлось побывать в начале и конце времен, там, где растворяется даже само понятие пространства. Что же мы в результате выяснили? Мы узнали, что нашим теориям пока еще недоступны ответы на эти вопросы — но что полезного это нам дает для понимания сущности разума? Я все же полагаю, что несмотря на отсутствие адекватной теории, мы можем извлечь важные уроки из нашего путешествия. А теперь нам необходимо вернуться домой. И хотя наше путешествие назад будет еще более наполнено догадками и предположениями, но других приемлемых путей назад просто нет!

Примечания

1. Некоторые «ревностные поборники» релятивизма могли бы предпочесть использовать световые конуса наблюдателей, а не их пространства одновременных событий. Однако, все сделанные нами заключения от этого не изменятся.
2. После первого просмотра напечатанного варианта мне вдруг пришло в голову, что оба человека должны были умереть задолго до этого. «Сопоставить свои наблюдения», в принципе, могли бы их *отдаленные потомки* (до которых вся информация о возникшем когда-то споре дошла бы, передаваясь из поколения в поколение).
3. Во внутризвездных процессах слияния легких ядер (например, водорода) в более тяжелые (например, гелий или в конечный продукт — железо) энтропия возрастает. По этой причине водород, присутствующий на Земле, часть которого мы можем, в конце концов, использовать путем его превращения в гелий на термоядерных станциях, содержит много «низкой энтропии». Возможность увеличения энтропии таким способом возникает только благодаря тому, что гравитация собрала ядра вместе, вдали от того гораздо большего числа фотонов, которые рассеялись по всему пространству и в настоящий момент образуют чернотельное фоновое излучение с температурой 2,7 К (см. с. 283). Это излучение заключает в себе существенно большую энтропию, чем та, которая содержится в веществе звезд и, если бы было возможно собрать это излучение и поместить его обратно в вещество звезд, то оно разложило бы большую часть тяжелых ядер на составляющие их более легкие ядра! Следовательно, прирост энтропии в процессе термоядерного синтеза является «временным» и возможен только благодаря концентрирующему воздействию гравитации. Позднее мы увидим, что, хотя энтропия, порождаемая в процессе термоядерного синтеза, намного превосходит энтропию, возникающую в большей части различных гравитационных процессов, а энтропия чернотельного фонового излучения оказывается еще большей — все это справедливо только временно и локально. Гравитационные запасы энтропии оказываются неизмеримо более мощными и существенно превосходят как энтропию термоядерного синтеза, так и энтропию фонового излучения.
4. Недавние результаты исследований сверхглубоких скважин на территории Швеции можно интерпретировать как подтверждающие теорию Голда, но ситуация далеко неоднозначна и может иметь альтернативное истолкование в рамках общепринятых геологических концепций.
5. Я предполагаю здесь, что эта звезда относится к так называемому «типу II» сверхновых. Если бы это была сверхновая «типа I», мы могли бы опять вести рассуждения в терминах «временного» прироста энтропии, связанного с термоядерным синтезом (см. примечание 3). Вряд ли, однако, сверхновая «типа I» способна произвести много урана.
6. Я отношу модели с нулевой и отрицательной пространственной кривизной к бесконечным моделям. Есть, однако, возможность некоторой «свертки» этих моделей, после которой они становятся пространственно конечными. Такое рассмотрение, которое вряд ли применимо к реальной вселенной,

существенно не меняет ход нашего обсуждения, и поэтому я предлагаю не заострять здесь внимание на этом вопросе.

7. Экспериментальные основания для такой уверенности заключаются, главным образом, в данных двух типов. Во-первых — это поведение частиц при их столкновениях друг с другом на различных скоростях: рассеяние, распад и рождение новых частиц. Эти процессы изучаются либо на ускорителях, построенных и размещенных в самых разных уголках Земли, либо с помощью космических лучей, бомбардирующих Землю из открытого космоса. Во-вторых, известно, что параметры, регулирующие взаимодействие частиц, не изменились даже на одну миллионную за 10^{10} лет (см. Барроу [1988]), так что, скорее всего, они существенно и не менялись (если менялись вообще) со времен первичного протошара.

8. На самом деле, принцип Паули не запрещает электронам находиться в одном и том же месте, а запрещает им находиться лишь в одном и том же «состоянии», учитывающем их движение и вращение (спин). Применение этого принципа в рассматриваемом случае связано с определенными тонкостями, и в первое время расценивалось некоторыми (особенно Эддингтоном) как достаточно спорное.

9. Похожие соображения были высказаны уже в 1784 году английским астрономом Джоном Мичеллом и, немного позднее и независимо от него, Лапласом. Они пришли к выводу, что наиболее массивные и плотные тела во вселенной могут оказаться совершенно невидимыми — как и черные дыры — но их (поистине пророческие) выводы были сделаны на основе ньютоновской теории, в которой подобные заключения являются, в лучшем случае, спорными. Надлежащий общерелятивистский подход был разработан Робертом Оппенгеймером и Хартландом Снайдером [1939].

10. На самом деле, точное положение горизонта (в общем случае нестационарной) черной дыры не может быть установлено непосредственными измерениями. В частности, для его определения необходимо обладать информацией о том веществе, которое черная дыра поглотит в будущем!

11. Смотри изложение этого вопроса в работах Белинского, Халатникова и Лифшица [1970] и Пенроуза [1979].

12. Возникает искушение отождествить гравитационный вклад в энтропию системы с некоторой мерой вейлевской кривизны, но до сих пор ни одной подходящей меры не найдено. (Искомая мера, вообще говоря, должна была бы обладать нелокальными свойствами.) К счастью, в наших рассуждениях мы можем обойтись и без нее.

13. Существует популярная в настоящее время точка зрения, называемая «инфляционным сценарием», которая призвана объяснить, почему вселенная, помимо всего прочего, является однородной на очень больших масштабах. Согласно этой теории, вселенная на очень ранних стадиях испытала гигантское расширение — намного превосходящее по своим масштабам «обычное» расширение стандартного сценария. Идея заключается в том, что любые нерегулярности сглаживаются в результате такого расширения. Однако, инфляция не работает без наложения еще более жестких начальных ограничений, чем те, которые уже даются гипотезой о вейлевской кривизне. Она не вводит в теорию никакой асимметричной во времени составляющей, которая дала бы возможность объяснить различие между начальной и конечной сингулярностью. (Более того, она опирается на физические теории — теории великого объединения — чей статус не более, чем ПРОБНЫЙ, по терминологии главы 5. Для более подробного знакомства с критическим анализом «инфляции», в контексте идей, изложенных в этой главе, см. Пенроуз [1989б].)

Глава 8

В поисках квантовой теории гравитации

Зачем нужна квантовая теория гравитации?

Что еще осталось узнать о мозге и мышлении такого, чего мы не выяснили в предыдущей главе? Хотя мы уже кратко рассмотрели некоторые из всеобъемлющих физических принципов, лежащих в основе направленности воспринимаемого нами «потока времени», нам все же пока не удалось понять не только почему мы воспринимаем время как нечто текущее, но даже почему мы вообще его воспринимаем. Я считаю, что тут необходимы гораздо более радикальные идеи. До сих пор мое изложение особым радикализмом не отличалось, хотя в некоторых случаях расстановка акцентов была далека от традиционной. Мы ознакомились со вторым началом термодинамики, и я попытался убедить читателя в том, что этот закон — данный нам в виде, выбранном самой природой, — уходит своими корнями в чрезвычайно сильное геометрическое ограничение на происхождение вселенной в результате Большого взрыва — *гипотезу о вейлевской кривизне*. Некоторые космологи предпочитают интерпретировать это исходное ограничение иначе, но такого рода ограничение на начальную сингулярность действительно является необходимым. *Выводы*, которые я собираюсь сделать из этой гипотезы, будут гораздо менее традиционными, чем сама гипотеза. Я утверждаю, что потребуются изменения в самих основах квантовой теории!

Эти изменения должны сыграть свою роль при объединении квантовой механики с общей теорией относительности, т. е. в рамках искомой *квантовой теории гравитации*. Большинство физиков не считают необходимым что-либо менять в квантовой теории при ее объединении с общей теорией относительности. Более того, они утверждают, что на пространственных масштабах, имеющих значение для нашего мозга, эффекты *любой* квантовой теории гравитации пренебрежимо малы! Они отмечают (и весьма резонно), что хотя такого рода физические эффекты действительно могут оказаться существенными на абсурдно малых пространственных масштабах, сравнимых с так называемой *планковской длиной* [1]), что составляет 10^{-35} м — т. е. примерно в 100 000 000 000 000 000 000 раз меньше размера самой маленькой из субатомных частиц, — эти эффекты тем не менее никоим образом напрямую не затрагивают явления, происходящие на много-много бóльших «обычных» пространственных масштабах, от 10^{-12} м и более, там, где правят бал химические и электрические процессы, важные для деятельности мозга. Собственно говоря, даже *классическая*

[1]) Это расстояние (10^{-35} м $= \sqrt{\hbar G c^{-3}}$), на котором так называемые «квантовые флуктуации» самой метрики пространства-времени становятся настолько большими, что обычное представление об однородном пространственно-временнóм континууме оказывается неприменимым. (Квантовые флуктуации являются следствием принципа неопределенности Гейзенберга — см. с. 223.)

(то есть неквантовая) теория гравитации почти никак не затрагивает эти электрические и химические процессы. Если классической гравитацией можно пренебречь, то какое может иметь значение любая ничтожно малая «квантовая поправка» к классической теории? Более того, поскольку *отклонения* от квантовой теории до сих пор не наблюдались, то *тем более* представляется лишенной всяких оснований сама мысль о каком бы то ни было влиянии на процессы мышления любого ничтожно малого гипотетического отклонения от стандартной квантовой теории!

Я же буду рассуждать совсем иначе. Меня интересует не столько влияние квантовой механики на теорию структуры пространства-времени (теорию относительности Эйнштейна), сколько возможное *обратное* влияние эйнштейновской теории пространства-времени на саму структуру квантовой механики. Я хочу подчеркнуть, что предлагаемая мною точка зрения *нетрадиционна*. Нетрадиционным является предположение о самой возможности влияния общей теории относительности на структуру квантовой механики! Традиционная физика относится с большим предубеждением к любым попыткам что-либо изменить в стандартной структуре квантовой механики. Несмотря на, по-видимому, непреодолимые трудности, возникающие при попытках непосредственного применения правил квантовой механики к теории Эйнштейна, работающие в этой области исследователи, как правило, делали отсюда вывод о необходимости корректировки *теории Эйнштейна*, а не квантовой механики [1]. Я же придерживаюсь практически противоположной точки зрения и считаю, что проблемы самой квантовой теории носят фундаментальный характер. Вспомним о несовместимости двух основных ее процедур — **U** и **R** (**U** подчиняется совершенно детерминистскому *уравнению Шредингера* — это так называемое уравнение *унитарной* эволюции, а **R** представляет собой вероятностную *редукцию вектора состояния*, необходимость в которой возникает всякий раз, когда предполагается, что было сделано «наблюдение»). По-моему, эту несовместимость *нельзя* адекватно разрешить простой подходящей «интерпретацией» квантовой механики (хотя эта точка как раз и является господствующей), — ее устранение возможно лишь в рамках новой теории, коренным образом отличной от существующей, в которой процедуры **U** и **R** будут рассматриваться как различные (и очень хорошие) приближения к более всеобъемлющей и точной *единой* процедуре. Моя точка зрения, следовательно, состоит в том, что даже такая изумительно точная теория, как квантовая механика, потребует изменений, и что именно теория относительности Эйнштейна позволит лучше всего понять характер этих изменений. Я пойду еще дальше, утверждая, что речь идет именно об искомой *квантовой теории гравитации*, одним из компонентов которой должна как раз стать предполагаемая единая процедура **U**/**R**.

С другой стороны, *с общепринятой точки зрения* любые прямые следствия квантовой теории гравитации должны иметь более эзотерический характер. Я уже упоминал об ожидаемом радикальном изменении структуры пространства-времени на абсурдно малых масштабах порядка планковской длины. Существует мнение (и, по-моему, вполне обоснованное), что квантовая теория гравитации должна сыграть фундаментальную роль в окончательном установлении природы наблюдаемого «зоопарка элементарных частиц». Например, сейчас у нас нет хорошей теории, которая бы объяснила, почему массы частиц именно таковы, каковы они есть — а ведь понятие «массы» теснейшим образом связано с понятием гравитации. (Действительно, единственное действие массы — быть «источником» гравитации.) К тому же не без оснований считается, что (согласно идее, выдвинутой где-то около 1955 года шведским физиком Оскаром Клейном) правильная квантовая теория гравитации обязана устранить расходимости, преследующие обычную квантовую теорию поля (см. с. 255). Физика представляет собой единое целое, и *правильная* квантовая теория гравитации, когда она, наконец, будет построена, должна стать основой нашего досконального понимания универсальных законов природы.

Мы, однако, пока еще далеки от такого понимания. Более того, вне всякого сомне-

ния любая гипотетическая квантовая теория гравитации не будет иметь практически никакого отношения к явлениям, управляющим поведением мозга. *Особенно* далеки от деятельности мозга могут оказаться те (общепринятые) аспекты квантовой теории гравитации, которые необходимы для выхода из тупика, в который мы попали в предыдущей главе, а именно для разрешения проблемы *пространственно-временных сингулярностей* — сингулярностей классической теории Эйнштейна, которые возникают в момент *большого взрыва* и в *черных дырах*, а также при *большом коллапсе* — если наша вселенная решит в конце концов сколлапсировать сама на себя. Конечно же, эта роль квантовой теории гравитации вполне может *показаться* далекой [от проблем деятельности мозга]. Я, однако, утверждаю, что тут все же имеется почти неуловимая, но важная логическая связь. Постараемся выяснить, в чем она состоит.

Что скрывается за гипотезой о вейлевской кривизне?

Как я уже отмечал, даже согласно традиционной точке зрения именно квантовая теория гравитации должна прийти на помощь классической общей теории относительности и решить проблему пространственно-временных сингулярностей. Так, квантовая теория гравитации должна дать непротиворечивое физическое описание взамен бессмысленного «бесконечного» результата классической теории. Я безусловно согласен с этой точкой зрения: это как раз та самая ситуация, где квантовая теория гравитации должна проявить себя в полной мере. Однако, теоретики не могут смириться с тем поразительным фактом, что проявления квантовой теории гравитации вопиющим образом асимметричны во времени! В случае Большого взрыва — *прошлой сингулярности* — квантовая теория гравитации должна требовать выполнения условия типа

$$\text{ВЕЙЛЬ} = 0$$

в тот момент, когда приобретает смысл описание в терминах классических понятий геометрии пространства-времени. С другой стороны, для сингулярностей, расположенных внутри черных дыр, и (возможно) для сингулярности большого коллапса — т. е. для *будущих сингулярностей* — такого рода ограничение отсутствует, и мы полагаем, что по мере приближения к такой сингулярности тензор Вейля стремится к бесконечности:

$$\text{ВЕЙЛЬ} \to \infty.$$

Я считаю это обстоятельство несомненным свидетельством асимметричности во времени искомой истинной теории. Итак:

искомая квантовая теория гравитации асимметрична во времени.

Хочу предупредить здесь читателя, что приведенный вывод, несмотря на его очевидность, с неизбежностью вытекающей из изложенных выше рассуждений, не является, тем не менее, общепринятым! Большинство исследователей, работающих в рассматриваемой области науки, крайне неохотно встают на эту точку зрения. Причина, по-видимому, кроется в отсутствии ясного способа, каким привычные и (насколько можно судить) хорошо нами понятые процедуры квантования могли бы породить асимметричную во времени [2] квантовую теорию при том, что классическая теория, к которой упомянутые процедуры применяются (стандартная общая теория относительности или ее модификации), сама по себе симметрична во времени. Соответственно, эти специалисты по квантованию гравитации вынуждены (если они вообще задаются подобными вопросами — что случается не так уж и часто) искать другие «объяснения» малого значения энтропии при Большом взрыве.

Многие физики могут возразить, что гипотезы, подобные предположению о нулевом начальном значении вейлевской кривизны, — представляя собой выбор «граничного условия», а не динамические законы, — находятся за пределами наших возможностей объяснения. Они утверждают, по сути, что в данном случае мы имеем дело с «актом Творца» и нечего даже и пытаться понять, почему нам дано именно это граничное условие, а не какое-нибудь другое. Однако, как мы уже убедились выше,

ограничение, накладываемое рассматриваемой гипотезой на «булавку Творца», по своей исключительности и точности не уступает той потрясающей и тончайшим образом организованной хореографии динамических законов, к пониманию которых мы пришли через уравнения Ньютона, Максвелла, Эйнштейна, Шредингера, Дирака и др. Хотя второе начало термодинамики и может показаться нечетким и статистическим по своей природе, оно тем не менее вытекает из чрезвычайно точного геометрического ограничения. Поскольку научное осмысление доказало свою ценность как способ понимания динамических уравнений, мне представляется неразумным впадать в отчаяние и терять всякую надежду на научное постижение ограничений, действовавших в случае «граничного условия», каким являлся Большой взрыв. С моей точки зрения, как одно, так и другое являются частью науки, хотя и той частью, которая нами — пока еще — недостаточно понята.

История науки продемонстрировала, насколько ценной для физики оказалась идея отделения *динамических уравнений* (законов Ньютона, уравнений Максвелла и т. д.) от так называемых *граничных условий* — то есть условий, необходимых для выделения из огромного множества решений того, что имеет физический смысл. Исторически простые формулировки были найдены именно для динамических уравнений. Движения частиц подчиняются простым законам, а вот о встречающихся во вселенной *реальных конфигурациях* частиц это, похоже, можно сказать нечасто. Иногда эти конфигурации на первый взгляд выглядят простыми — как, например, в случае планетных орбит, эллиптическая форма которых была установлена Кеплером, — но простота их в дальнейшем оказалась *следствием* динамических законов. Более глубокое понимание всегда достигалось через динамические законы, а простые конфигурации, подобные вышеописанной, как правило оказывались просто приближениями к более сложным конфигурациям вроде возмущенных (уже не совсем эллиптических) реально наблюдаемых движений планет, которые находят свое объяснение в динамических уравнениях Ньютона. Граничные условия служат для «запуска» рассматриваемой системы, после чего за дело принимаются динамические законы. Сам факт возможности отделения проблемы динамического поведения от вопроса о конфигурации реального содержимого вселенной представляет собой одно из важнейших достижений физической науки.

Я сказал, что исторически это разделение на динамические уравнения и граничные условия сыграло чрезвычайно важную роль. Сама же возможность такого разделения представляет собой свойство *конкретного* типа уравнений (дифференциальных уравнений), который, как кажется, всегда возникает в физике. Но я не верю, что это разделение сохранится навечно. По-моему, когда нам удастся окончательно постичь законы или принципы, в *действительности* управляющие поведением нашей вселенной, — а не просто те изумительные приближения, к пониманию которых мы уже пришли и которые суть составные части наших ПРЕВОСХОДНЫХ современных теорий, то увидим, как различие между динамическими уравнениями и граничными условиями исчезнет, уступив место потрясающе согласованной всеобъемлющей схеме. Разумеется, утверждая это, я выражаю исключительно свое собственное мнение, с которым многие могут не согласиться. Но именно эту точку зрения я имею в виду, когда стараюсь нащупать следствия из пока неизвестной квантовой теории гравитации. (Под этим углом будут рассмотрены также некоторые наиболее спекулятивные рассуждения последней главы.)

Как же можно изучать следствия неизвестной еще теории? Это, однако, не обязательно столь безнадежно, как кажется. Главное здесь — быть последовательными! Сначала я попрошу вас допустить, что наша гипотетическая теория, далее называемая ПКТГ («правильная квантовая теория гравитации»), должна объяснить гипотезу о вейлевской кривизне (ГВК). Это значит, что в непосредственном ближайшем будущем *начальная* сингулярность должна удовлетворять условию ВЕЙЛЬ = 0. Это ограничение не должно противоречить законам ПКТГ и поэтому обязано соблюдаться для *любой* начальной сингулярности, а не только той,

что мы называем Большим взрывом. При этом я никоим образом не утверждаю существование в нашей реальной вселенной каких бы то ни было других начальных сингулярностей, отличных от Большого взрыва, но всего лишь говорю, что такая сингулярность, *если* бы она существовала, должна удовлетворять ограничению, накладываемому ГВК. Начальная сингулярность — это сингулярность, из которой, в принципе, могут возникать частицы. Такие сингулярности ведут себя противоположно черным дырам — *конечным* сингулярностям, в которые частицы могут падать.

Одним из возможных примеров начальной сингулярности, отличной от Большого взрыва, может быть сингулярность *белой дыры*, которая, как мы помним из главы 7, представляет собой обращенную во времени черную дыру (см. рис. 7.14). Но, как мы уже видели, сингулярности внутри черных дыр удовлетворяют условию ВЕЙЛЬ $\to \infty$, поэтому и для белых дыр должно выполняться ВЕЙЛЬ $\to \infty$. Однако теперь сингулярность стала начальной и для нее, согласно ГВК, должно выполняться условие ВЕЙЛЬ $= 0$. Таким образом, ГВК делает существование белых дыр в нашей вселенной *невозможным*! (К счастью, этот результат не только желателен из термодинамических соображений — поскольку белые дыры были бы вопиющим нарушением второго начала термодинамики — но к тому же согласуется с наблюдательными данными! Время от времени разными астрофизиками предпринимались попытки объяснить некоторые явления, предполагая существование белых дыр, но такие гипотезы всегда создавали гораздо больше проблем, чем решали.) Заметьте, что сам Большой взрыв я не называю «белой дырой». Белая дыра должна содержать *локализированную* начальную сингулярность, для которой выполнение условия ВЕЙЛЬ $= 0$ невозможно, в то время как всеобъемлющий Большой взрыв *может* удовлетворять условию ВЕЙЛЬ $= 0$, и существование такого взрыва допускается ГВК при условии выполнения соответствующего ограничения.

Примером еще одного вида начальных сингулярностей является *точка взрыва черной дыры*, окончательно *исчезающей*, после, скажем, 10^{64} лет хокинговского испарения (см. с. 298, а также с. 314)! Точная природа этого (весьма правдоподобно аргументированного) явления является предметом многочисленных теоретических гипотез. Я думаю, что никакого противоречия с ГВК здесь нет. Такого рода (локализированный) взрыв может быть практически мгновенным и симметричным, и я не вижу здесь никакого конфликта с гипотезой ВЕЙЛЬ $= 0$. Во всяком случае, если предположить, что черных мини-дыр не существует (см. с. 298), то первый такой взрыв вряд ли произойдет раньше, чем вселенная просуществует в 10^{54} раз больше современного возраста T. Чтобы получить представление о величине $10^{54} \times T$, мысленно уменьшим T до самого короткого измеримого промежутка времени, равного времени распада самой короткоживущей из нестабильных частиц. В полученной таким образом шкале времени *современный* возраст вселенной окажется меньше $10^{54} \times T$ в миллион миллионов раз!

Кто-нибудь может посмотреть на все это с другой точки зрения. Мне могут возразить [3], что ПКТГ не обязана быть асимметричной во времени, а должна лишь допускать на самом деле два типа сингулярностей, для одних из которых должно выполняться равенство ВЕЙЛЬ $= 0$, а для вторых возможно ВЕЙЛЬ $\to \infty$. В нашей вселенной оказалась сингулярность первого типа, и наше восприятие направления течения времени (в силу вытекающего отсюда второго начала термодинамики) помещает эту сингулярность туда, где находится наше так называемое «прошлое», а не «будущее». По-моему, однако, соображение это в таком виде не выдерживает критики. Оно не объясняет отсутствие *других* начальных сингулярностей типа ВЕЙЛЬ $\to \infty$ (а также отсутствие других начальных сингулярностей типа ВЕЙЛЬ $= 0$). Почему, если согласиться с этой точкой зрения, вселенная не усеяна белыми дырами? Поскольку она, как мы предполагаем, *кишит черными дырами*, отсутствие белых дыр требует объяснения [2].

---

[2] Некоторые могут на это возразить (совершенно справедливо), что наблюдения не подтверждают однозначным образом мое утверждение о существовании во вселенной черных дыр и отсутствии белых. Но мой довод, в основном, теоретического характера. Черные дыры не противоречат второму началу термодинамики,

Другое соображение, иногда привлекаемое в связи с рассматриваемой проблемой, — это так называемый *антропный принцип* (см. Барроу, Типлер [1986]). Согласно этому соображению, конкретная вселенная, обитателями которой мы сейчас являемся, выбрана из всех *возможных* вселенных потому, что в ней должны существовать мы (или, по крайней мере какие-нибудь чувствующие существа), чтобы ее было кому наблюдать! (Я вернусь к обсуждению антропного принципа в главе 10.) На этом основании утверждается, что разумные существа могут населять только вселенные с Большим взрывом очень определенного типа — и поэтому следствием этого принципа должно быть что-то вроде ГВК. Однако, это соображение не позволяет и близко подойти к числу $10^{10^{123}}$, полученному в главе 7 (см. с. 300), которое характеризует степень «специфичности» Большого взрыва. Путем очень грубого расчета можно установить, что порождение солнечной системы со всем ее населением в результате случайных столкновений частиц обойдется гораздо «дешевле», а именно: соответствующая степень «невероятности» (измеряемая в терминах фазовых объемов) соответствует «всего лишь» одной доле из много менее чем $10^{10^{60}}$. Это все, что может дать антропный принцип, и нам еще чудовищно далеко до требуемого числа. Более того, соображения, основанные на антропном принципе, не в состоянии объяснить, как и обсуждавшаяся перед этим концепция, отсутствия белых дыр.

Временна́я асимметрия в редукции вектора состояния

По-видимому, нам действительно ничего не остается, как заключить, что ПКТГ должна быть асимметричной во времени теорией, одним из следствий которой является ГВК (или что-то вроде этого). Как же асимметричная во времени теория может получиться из симметричных во времени ингредиентов: квантовой теории и общей теории относительности? Есть, оказывается, несколько технических способов достижения этой цели, и ни один из них не исследовался достаточно глубоко (см. Аштекар и др. [1989]). Но я собираюсь подойти к проблеме с другой стороны. Как я уже отмечал, квантовая теория «симметрична во времени», но это в действительности относится только к части U теории (уравнению Шредингера и т. д.). Обсуждая временну́ю симметрию физических законов в начале главы 7, я умышленно избегал упоминания части R (коллапс волновой функции). Согласно преобладающей точке зрения R тоже должна быть, по-видимому, симметричной во времени. Своим существованием эта точка зрения может, в частности, быть обязана нежеланию признавать в R реальный независимый от U «процесс», вследствие чего из временно́й симметрии U должна бы также вытекать временна́я симметрия R. Я хотел бы возразить, что это *не так*: R *асимметрична* во времени — по крайней мере, если считать R просто процедурой, принятой физиками для расчета квантово-механических вероятностей.

Я сначала напомню вам используемую в квантовой механике так называемую процедуру редукции вектора состояния (R) (см. рис. 6.23). Рис. 8.1 иллюстрирует (условно) характер предполагаемой эволюции вектора состояния $|\psi\rangle$ в квантовой механике. Как видим, этот характер довольно своеобразный: считается, что большую часть времени эволюция происходит в соответствии с *унитарной* эволюционной процедурой U (уравнение Шредингера), но в некоторые моменты времени, когда предполагается, что происходит «наблюдение» (или «измерение»),

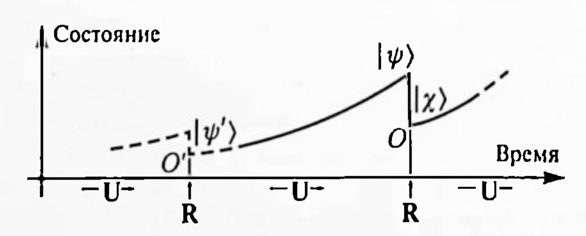


Рис. 8.1. Временная эволюция вектора состояния: гладкая унитарная эволюция U (в соответствии с уравнением Шредингера), перемежаемая с разрывной редукцией R вектора состояния

а белые дыры противоречат! (Разумеется, можно просто постулировать второе начало термодинамики и отсутствие белых дыр, но мы хотим достичь более глубокого понимания сути вещей, происхождения второго начала термодинамики.)

применяется **R**-процедура и вектор состояния *скачком* переходит в другой вектор состояния, например $|\chi\rangle$, где $|\chi\rangle$ представляет собой одну из двух или нескольких ортогональных альтернативных возможностей $|\chi\rangle$, $|\varphi\rangle$, $|\theta\rangle$, ..., определяемых природой конкретного производимого наблюдения O. Тогда вероятность p скачкообразного перехода от $|\psi\rangle$ к $|\chi\rangle$ определяется уменьшением квадрата длины $|\psi|^2$ вектора $|\psi\rangle$ при проекции $|\psi\rangle$ (в гильбертовом пространстве) на направление вектора $|\chi\rangle$. (Математически это равно величине уменьшения $|\chi|^2$ при проекции вектора $|\chi\rangle$ на направление $|\psi\rangle$.) В таком виде эта процедура оказывается асимметричной во времени, поскольку сразу же *после* выполнения наблюдения O вектор состояния должен принадлежать к *заданному множеству* $|\chi\rangle$, $|\varphi\rangle$, $|\theta\rangle$, ... возможных значений, *определяемых* O, в то время как непосредственно *перед* наблюдением O вектор состояния должен был иметь значение $|\psi\rangle$, которое не обязано быть равным ни одному из элементов упомянутого множества. Однако, это всего лишь кажущаяся асимметричность и она может быть устранена, если посмотреть на эволюцию вектора состояния с другой точки зрения. Рассмотрим квантово-механическое решение, *обращенное во времени*. Это экстравагантное описание проиллюстрировано на рис. 8.2. Мы предполагаем, что вектор состояния равен $|\chi\rangle$ непосредственно *перед* O, а не сразу после этого наблюдения, и применим процедуру унитарной эволюции вспять по времени вплоть до момента *предыдущего* наблюдения O'. Предположим, что в результате обратной эволюции мы получим состояние, описываемое вектором $|\chi'\rangle$ (сразу же после наблюдения O'). В нормальном описании эволюции вперед во времени, изображенном на рис. 8.1, сразу же вслед за O' мы имели другое состояние $|\psi'\rangle$ (результат наблюдения O', при котором эволюция вперед во времени вектора $|\psi'\rangle$ переводит его в $|\psi\rangle$ в момент наблюдения O). Теперь в нашем обращенном во времени описании у вектора $|\psi'\rangle$ тоже есть своя роль: он представляет состояние системы непосредственно перед O'. Вектор состояния $|\psi'\rangle$ соответствует состоянию, фактически наблюдавшемуся в точке O', так что с «обращенной»

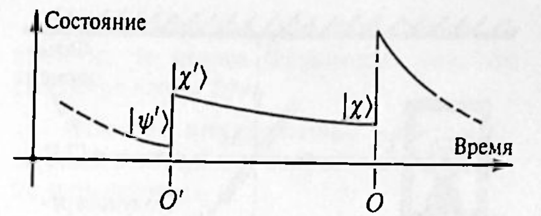


Рис. 8.2. Более экстравагантное изображение эволюции вектора состояния, описанное вспять по времени. Расчетная вероятность, связывающая наблюдение в точке O с наблюдением в точке O', такая же, как и в случае, изображенном на рис. 8.1, но к чему относится это вычисленное значение?

точки зрения мы рассматриваем $|\psi'\rangle$ как результат наблюдения O' в *обращенном вспять времени*. Расчетное значение квантово-механической вероятности p', связывающее результаты наблюдений в точках O и O', теперь определяется уменьшением величины $|\chi'|^2$ при проекции $|\chi'\rangle$ в направлении $|\psi'\rangle$ (что равно уменьшению $|\psi'|^2$ при проекции $|\psi'\rangle$ в направлении $|\chi'\rangle$). То, что мы получим то же самое значение, что и раньше, является фундаментальным свойством оператора **U** [4].

Таким образом, *может создаться видимость* установления *симметричности во времени квантовой теории* даже в случае, когда помимо обычной процедуры унитарной эволюции **U** учитывается также и разрывный процесс, описываемый процедурой редукции **R** вектора состояния. Это, однако, *неверно*. Квантовая вероятность p описывает — независимо от того, как она рассчитывается — вероятность получить результат (а именно, $|\chi\rangle$) в точке O при условии определенного результата (а именно, $|\psi'\rangle$) в точке O'. Эта вероятность не обязательно равна вероятности получить данный результат в точке O' *при условии* данного результата в точке O, а ведь именно последнюю вероятность [5] и должна определить обращенная во времени квантовая механика. Просто удивительно, до чего много физиков молчаливо полагают эти две вероятности равными друг другу. (Я сам этим грешил — см. Пенроуз [1979б], с. 584.) Однако наиболее вероятно, что эти две вероятности совершенно различны и только первая из них правильно определяется в рамках квантовой механики!

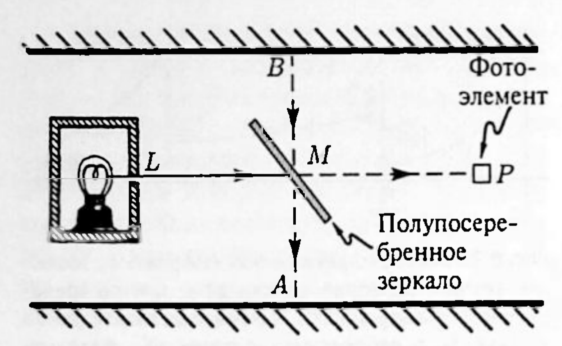


Рис. 8.3. Необратимость во времени R-процедуры в простом квантовом эксперименте. Вероятность регистрации фотона фотоэлементом *при условии* излучения фотона источником равна в точности одной второй, но вероятность излучения фотона источником *при условии*, что фотоэлемент зарегистрировал фотон, заведомо не равна одной второй

Давайте поясним эту ситуацию на простом конкретном примере. Предположим, что у нас есть лампа L и фотоэлемент (то есть, детектор фотонов) P. Между L и P разместим полупосеребренное зеркало M, наклонив его под углом равным, скажем, 45° к линии, соединяющей точки L и P (рис. 8.3). Предположим, что лампа время от времени случайным образом испускает фотоны, и что конструкция ее такова (в ней используются параболические зеркала), что фотоны всегда оказываются очень точно нацеленными на P. При каждом попадании фотона на фотоэлемент последний регистрирует это событие, причем мы предполагаем, что устройство срабатывает со 100 %-ной надежностью. Предположим также, что каждый факт излучения фотона регистрируется в точке L и тоже со 100 %-ной надежностью. (Ни одно из этих идеализированных требований не противоречит принципам квантовой механики, хотя практическое достижение такой эффективности может представлять определенные трудности.)

Свойства полупосеребренного зеркала M таковы, что оно отражает в точности половину попадающих на него фотонов и пропускает остальную половину. Правильнее рассматривать это с точки зрения квантовой механики. Волновая функция фотона падает на зеркало и расщепляется на две волновых функции. Амплитуда отраженной части волны равна $1/\sqrt{2}$, а амплитуда прошедшей части волны тоже равна $1/\sqrt{2}$. Обе части волновой функции должны считаться «сосуществующими» (при нормальном описании вперед по времени) до того момента, когда предполагается имевшим место «наблюдение». В этой точке ситуация с одновременно сосуществующими альтернативами разрешается (в пользу одной *или* другой) *фактически реализованной* альтернативы с вероятностями, равными квадратам (модулей) соответствующих амплитуд, а именно $(1/\sqrt{2})^2 = 1/2$ в обоих случаях. После выполнения наблюдения вероятности отражения или прохождения фотона действительно оказываются равными одной второй.

Посмотрим теперь, как все это соотносится с нашим экспериментом. Предположим, что зарегистрирован факт излучения фотона лампой L. Волновая функция фотона расщепляется на зеркале и приходит в точку P с амплитудой, равной $1/\sqrt{2}$, поэтому фотоэлемент либо регистрирует фотон, либо не регистрирует его — и то и другое с вероятностью, равной одной второй. Другая часть волновой функции фотона попадает в точку A на *лабораторной стене* (см. рис. 8.3) и тоже с амплитудой $1/\sqrt{2}$. Если фотоэлемент P *не* регистрирует событие, то фотон следует считать попавшим в лабораторную стену в точке A. Если бы в точке A находился другой фотоэлемент, то он регистрировал бы фотон всякий раз, когда фотоэлемент P не регистрирует фотон, и не регистрировал бы фотон всякий раз, когда фотоэлемент регистрирует фотон. В этом смысле нет никакой необходимости устанавливать фотоэлемент в точке A. Мы можем определить, что сделал *бы* фотоэлемент в точке A, будь он там установлен, просто глядя на фотоэлементы в точках L и P.

Теперь должно стать ясно, как выполняются расчеты в квантовой механике. Зададимся вопросом:

> «Если известно, что лампа L сработала, то какова вероятность того, что сработал фотоэлемент P?»

Для ответа на этот вопрос учтем, что имеется амплитуда, равная $1/\sqrt{2}$ для фотона, прошедшего путь LMP, и амплитуда, равная $1/\sqrt{2}$, для фотона, прошедшего

путь LMA. Возведя эти амплитуды в квадрат, получаем соответствующие вероятности, равные 1/2 и 1/2, попадания фотона в точки P и A соответственно. Следовательно, на наш вопрос квантовая механика дает ответ, равный

«одной второй».

И действительно, именно такой результат получился бы в случае проведения реального эксперимента.

Мы могли бы с таким же успехом использовать экстравагантную процедуру «с обращенным вспять временем» и получили бы тот же самый результат. Предположим, что мы зафиксировали факт срабатывания фотоэлемента в точке P. Рассмотрим направленную вспять во времени волновую функцию фотона в предположении, что фотон в конце концов приходит в точку P. Отслеживая эволюцию процесса назад во времени, мы видим, что фотон движется назад от P, пока не достигнет зеркала M. В этой точке происходит бифуркация волновой функции и мы имеем амплитуду $1/\sqrt{2}$ того, что фотон достигнет лампы L, и амплитуда $1/\sqrt{2}$ того, что фотон претерпит отражение в точке M и придет в *другую* точку на лабораторной стене, а именно в точку B на рис. 8.3. Возводя соответствующие амплитуды в квадрат, мы снова получаем для обеих вероятностей значения, равные одной второй. Следует, однако, отдавать себе отчет в том, на какие именно вопросы отвечают эти вероятности. А вопросы следующие: «Если известно, что лампа L сработала, то какова вероятность срабатывания фотоэлемента P?» — тот же самый вопрос, что мы рассматривали до этого; и более экстравагантный вопрос: «Какова вероятность срабатывания фотоэлемента P при условии, что известен факт испускания фотона из стены в точке B?»

Мы можем рассматривать оба ответа как экспериментально «правильные» в определенном смысле, хотя второй ответ (испускание фотона из стены) скорее представляет собой логическое умозаключение, а не результат *реально выполненного* ряда экспериментов! Однако ни один из этих вопросов не является *обращением во времени* того, что был задан выше. Обращенный вспять во времени вопрос звучал бы так:

«Если известно, что фотоэлемент P сработал, то какова вероятность того, что сработала лампа L?»

Отметим, что *правильный* экспериментальный ответ на этот вопрос — это никакая не «одна вторая», а

«единица».

В случае срабатывания фотоэлемента нет практически никаких сомнений в том, что фотон пришел от лампы, а не от лабораторной стены! На наш обращенный во времени вопрос проведенный в рамках квантовой механики расчет дал нам *абсолютно неверный ответ*!

Отсюда следует, что правила **R**-части квантовой механики просто-напросто неприменимы к такого рода обращенным во времени задачам. Если мы хотим рассчитать вероятность прошлого состояния исходя из известного состояния в будущем, то применение стандартной **R**-процедуры, которая заключается в простом возведении в квадрат модуля квантово-механической амплитуды, приводит к неверным результатам. Эта процедура пригодна только для расчета вероятностей *будущих* событий исходя из *прошлых* событий — и в этом случае она работает великолепно! Поэтому я считаю совершенно очевидным, что **R**-процедура *не может быть симметрична во времени* (и, между прочим, вследствие этого не выводима из симметричной во времени процедуры **U**).

Многие могут посчитать, что причина этого противоречия с временно́й симметрией состоит в том, что второму началу термодинамики каким-то образом все же удалось пролезть в цепь рассуждений и привнести дополнительную асимметрию во времени, не описываемую процедурой возведения амплитуды в квадрат. Действительно, кажется, что любой физический измерительный прибор, способный реализовать **R**-процедуру, должен содержать элемент «термодинамической необратимости» — так, что энтропия возрастает всякий раз, когда имеет место измерение. Я думаю, что процесс измерения должен быть фундаментальным образом связан со вторым началом термодинамики. Более того, по-видимому попытки обратить вспять во времени *целиком весь*

процесс квантово-механического эксперимента, вроде описанного выше (идеализированного) опыта, с регистрацией всех проведенных измерений, бессмысленны. Я не задавался вопросом о том, как далеко мы можем пойти по пути действительного обращения эксперимента во времени. Меня интересовала только применимость этой замечательной квантово-механической процедуры, которая дает правильные вероятности через вычисление квадратов модулей амплитуд. Поразительно, что эта простая процедура применима в направлении от прошлого к будущему и при этом не требует никакой дополнительной информации о системе. Действительно, *невозможность* повлиять на эти вероятности, которые в квантовой теории являются абсолютно *случайными*, представляет собой одну из неотъемлемых частей рассматриваемой теории! Однако попытка применить те же самые процедуры в направлении от будущего к прошлому (т. е. не для предсказания будущего, а для установления прошлого) приводит к результату неверному до удивления. Можно приводить сколько угодно оправданий, смягчающих обстоятельств и других доводов для объяснения того, *почему* процедура возведения амплитуды в квадрат не дает правильных результатов в случае применения ее в направлении от будущего к прошлому, но факт остается фактом. А при рассмотрении ситуации от прошлого к будущему никакие оправдания попросту не нужны! Процедура **R**, *в том виде, как она реально применяется*, просто-напросто не является симметричной во времени.

Ящик Хокинга: связь с гипотезой о вейлевской кривизне?

Как бы то ни было, а читатель вне всякого сомнения подумает: какое все это имеет отношение к ГВК или ПКТГ? Действительно, *второе начало термодинамики*, в его настоящем виде, вполне может быть частью процедуры **R**, но вот где тут в этих непрерывных «каждодневных» актах редукции вектора состояния может найтись место сколь-нибудь заметным эффектам пространственно-временных сингулярностей? Чтобы прояснить этот вопрос, я хочу, хотя и с совершенно иной целью, описать здесь фантастический «мысленный эксперимент», первоначально предложенный Стивеном Хокингом.

Представьте себе герметичный ящик чудовищных размеров. Его стенки предполагаются абсолютно отражающими и непроницаемыми для любого воздействия. Сквозь них не может пройти никакой материальный объект, в том числе никакой электромагнитный сигнал, нейтрино и вообще все что угодно. Стенки отражают обратно любой объект, независимо от того, приходит ли он снаружи или изнутри, и даже действие гравитации не может проникнуть сквозь них. Такие стенки невозможно сделать ни из одного существующего в природе вещества. Никто в действительности не в состоянии *выполнить* описанный ниже «эксперимент». (И, как мы увидим, никто и не захочет этого делать!) Важно не это. Целью мысленного эксперимента является раскрытие общих принципов путем простого мысленного рассмотрения *в принципе выполнимых* опытов. Технические проблемы игнорируются при условии, что они не связаны с рассматриваемыми общими принципами. (Вспомним дискуссию о шредингеровской кошке в главе 6.) В нашем случае проблемы сооружения стенок ящика должны рассматриваться с точки зрения стоящих перед нами целей как чисто «технические», и, следовательно, ими надо пренебречь.

Внутри ящика находится большое количество вещества. Для нас не имеет значения, что это за вещество. Нас интересует только его полная масса M, которая должна быть очень большой, а также большой объем V, в который она заключена. Что же мы собираемся делать с этим дорогостоящим ящиком, а также с его совершенно неинтересным содержимым? Мы произведем самый занудный из опытов, какой только можно себе вообразить. Оставим ящик в покое — навечно!

Нас интересует окончательная судьба того, что находится внутри. Согласно второму началу термодинамики, энтропия содержимого ящика должна возрастать, пока не достигнет максимума, а вещество — состояния «теплового равновесия». После

этого уже не будет происходить практически ничего интересного, если не считать «флуктуаций», приводящих к (относительно) кратковременным отклонениям от теплового равновесия. В нашей ситуации мы полагаем M достаточно большим при соответствующем V (т. е. *очень* большом, но не слишком большом), так что к моменту достижения «теплового равновесия» большая часть вещества сколлапсирует в *черную дыру*, окруженную совсем небольшим количеством вещества, и излучения, которые образуют (очень холодную!) тепловую ванну с погруженной в нее черной дырой. Чтобы быть конкретнее, примем M равной массе Солнечной системы, а V — размеру Млечного Пути! В этом случае температура «ванны» составит всего 10^{-7} градуса выше абсолютного нуля!

Чтобы лучше понять природу описываемых здесь равновесия и флуктуаций, вспомним понятие *фазового пространства*, с которым мы познакомились в главах 5 и 7, в частности, в связи с понятием энтропии. На рис. 8.4 условно изображено все фазовое пространство ℙ содержимого ящика Хокинга. Как мы помним, фазовое пространство — это пространство с большим количеством измерений, каждая точка которого полностью отображает одно из возможных состояний рассматриваемой системы — в данном случае содержимого ящика. Таким образом, каждая точка ℙ содержит информацию о положениях и импульсах всех находящихся в ящике частиц, а также всю необходимую информацию о *геометрии пространства-времени* внутри ящика. Расположенная в правой части рис. 8.4 подобласть 𝔹 (фазового пространства ℙ) представляет совокупность всех состояний с *черной дырой* внутри ящика (включая все случаи наличия более чем одной черной дыры), а расположенная слева область 𝔸 представляет совокупность всех состояний без черных дыр. Представим себе дальнейшее разбиение областей 𝔸 и 𝔹 на меньшие ячейки для построения «грубого разбиения», необходимого для точного определения энтропии (см. рис. 7.3 на с. 274). Точный вид этого разбиения нас здесь не интересует. На этом этапе нам важно лишь, что самая большая из рассматриваемых ячеек — та, что представляет со-

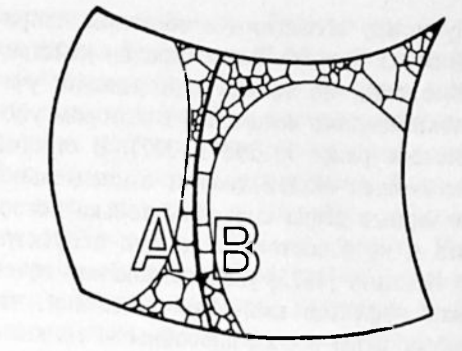


Рис. 8.4. Фазовое пространство ℙ ящика Хокинга. Область 𝔸 соответствует состояниям без черных дыр внутри ящика, а область 𝔹 — состояниям, при которых внутри ящика есть хотя бы одна черная дыра

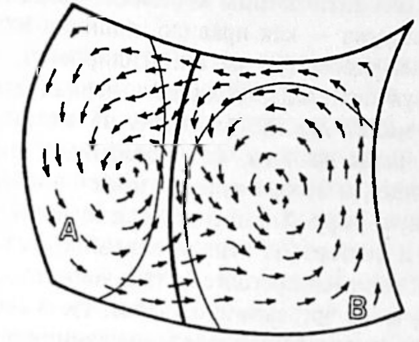


Рис. 8.5. «Гамильтонов поток» содержимого хокинговского ящика (см. рис. 5.11). Линии тока, пересекающие границу между областями в направлении от 𝔸 к 𝔹, соответствуют коллапсу в черную дыру, а линии, пересекающие границу от 𝔹 к 𝔸 — исчезновению черной дыры в результате хокинговского испарения

стояния теплового равновесия при наличии черной дыры, — занимает большую часть области 𝔹, а (несколько меньшая) большая часть области 𝔸 представляет то, что, *как кажется*, является тепловым равновесием, но без единой черной дыры.

Вспомним теперь, что на каждом фазовом пространстве существует поле стрелок (векторное поле), описывающих эволюцию физической системы во времени (см. главу 5, с. 170, а также рис. 5.11). Таким образом, чтобы узнать, что произойдет с нашей системой в следующий момент, нужно просто сдвинуться вдоль стрелок (рис. 8.5). Некоторые стрелки перейдут из области 𝔸 в область 𝔹. Такое происходит при возникновении черной дыры в результате гравитационного коллапса вещества. А пересекают ли какие-нибудь стрелки гра-

ницу между областями в обратном направлении из 𝔹 в 𝔸? Такие стрелки действительно есть, но только при условии учета *хокинговского испарения*, о котором upоминалось ранее (с. 298 и 307). В строгой *классической* общей теории относительности черные дыры способны только поглощать и не в состоянии ничего испускать. Но Хокингу [1975] удалось показать путем учета эффектов квантовой механики, что черные дыры все же способны — на квантовом уровне — кое-что испускать в процессе *хокинговского излучения*. (Это происходит в рамках квантового процесса «рождения виртуальных пар», при котором частицы и античастицы постоянно создаются из вакуума — как правило, лишь на мгновение, чтобы тут же аннигилировать, исчезнув без следа. Если есть черная дыра, она может «проглотить» одну из частиц такой пары до того, как произойдет аннигиляция, и вторая частица может покинуть черную дыру. Хокинговское излучение как раз и состоит из этих убежавших частиц.) При обычных обстоятельствах хокинговское излучение чрезвычайно слабое. Но в состоянии теплового равновесия величина энергии, теряемой черной дырой в результате хокинговского излучения, в точности компенсируется энергией, получаемой черной дырой в результате поглощения других «тепловых частиц» из окружающей «тепловой ванны», в которой дыра находится. В результате «флуктуаций» иногда может возникать небольшой избыток излучения или недостаток поглощения, что приводит к потере энергии черной дырой. Теряя энергию, черная дыра теряет также и массу (согласно формуле Эйнштейна $E = mc^2$) и, согласно законам, управляющим хокинговским излучением, становится чуть-чуть горячее. В очень редких случаях, если флуктуация оказывается достаточно большой, черная дыра может даже пойти в разнос, постоянно разогреваясь, теряя все больше энергии в этом процессе, непрерывно уменьшаясь в размерах, пока наконец (как мы предполагаем) совершенно не исчезнет в результате бурного взрыва! Когда это случится (и если считать, что других дыр в ящике нет), мы оказываемся в ситуации перехода из области 𝔹 в область 𝔸 фазового пространстве ℙ, и значит действительно есть стрелки, идущие из области 𝔹 в область 𝔸!

Я хотел бы сделать замечание о смысле, который я вкладываю здесь в понятие «флуктуация». Вспомним ячейки грубого разбиения, рассмотренные в предыдущей главе. Точки фазового пространства, принадлежащие одной ячейке, считаются (макроскопически) «неотличимыми» друг от друга. Энтропия возрастает, потому что, следуя вдоль стрелок, с течением времени мы, как правило, переходим ко все более крупным ячейкам. В конечном итоге точка фазового пространства оказывается затерянной внутри самой большой ячейки — а именно той, что соответствует тепловому равновесию (максимальной энтропии). Однако, это будет справедливо только до определенной степени. Если подождать достаточно долго, то точка фазового пространства окажется *в какой-то момент* в ячейке меньших размеров, и энтропия, соответственно, уменьшится. Как правило, это состояние продлится (сравнительно) недолго и энтропия вскоре снова увеличится при возвращении точки фазового пространства в самую крупную ячейку. Это — *флуктуация*, сопровождаемая мимолетным понижением энтропии. Обычно значительного падения энтропии не происходит, но в очень редких случаях возникает *огромная* флуктуация и энтропия может уменьшиться существенно и остаться малой на протяжении значительного времени.

Как раз такого рода событие и должно произойти, чтобы произошел переход из области 𝔹 в область 𝔸 через процесс хокинговского испарения. Очень большая флуктуация нужна потому, что маленькую ячейку необходимо протащить через то самое место, где стрелки пересекают границу между областями 𝔹 и 𝔸. Точно также, если наша точка фазового пространства находится внутри большой ячейки в области 𝔸 (представляющей совокупность состояний теплового равновесия без черных дыр), пройдет еще очень много времени, прежде чем произойдет гравитационный коллапс и точка перейдет внутрь области 𝔹. И снова нужна большая флуктуация. (Тепловое излучение неохотно идет на гравитационный коллапс!)

Каких стрелок больше — тех, что идут из 𝔸 в 𝔹; тех, что идут из 𝔹 в 𝔸; или же их число стрелок обоих типов *одинаково*? Для нас это очень важно. Вопрос можно сформулировать иначе: что природе «проще сделать» — породить черную дыру, заставив сколлапсировать частицы в состоянии теплового равновесия или же избавиться от черной дыры через хокинговское испарение? А может оба процесса одинаково «трудные»? Строго говоря, нас интересует не число стрелок, а скорость потока объема фазового пространства. Представьте себе, что фазовое пространство заполнено некой (многомерной) несжимаемой жидкостью. Стрелки отображают поток этой жидкости. Вспомним теперь описанную в главе 5 (с. 172) *теорему Лиувилля*, гласящую, что фазовый поток сохраняет объем элемента фазового пространства — а это как раз и означает, что наша жидкость, заполняющая фазовое пространство, действительно является несжимаемой! Теорема Лиувилля как будто говорит нам, что поток из 𝔸 в 𝔹 должен *равняться* потоку из 𝔹 в 𝔸, поскольку фазовая жидкость, будучи несжимаемой, не может накапливаться на одной какой-нибудь стороне. Таким образом, кажется, что черную дыру так же трудно создать из теплового излучения, как и разрушить ее!

К такому же выводу пришел и Хокинг, правда на основании несколько иных соображений. Главным аргументом Хокинга была *симметричность во времени* всех основных физических законов, имеющих отношение к рассматриваемой задаче (общая теория относительности, термодинамика, стандартные процедуры квантовой теории), из которой следует, что, если повернуть время вспять, то мы получим тот же самый результат, что и для прямого течения времени. Все сводится к простой смене направления всех стрелок в ℙ на противоположное. Из *этого* рассуждения действительно совершенно строго следует точное равенство числа стрелок из 𝔸 в 𝔹 числу стрелок из 𝔹 в 𝔸 *при условии, что* при обращении направления времени область 𝔹 отображается сама на себя (а область 𝔸 тоже, соответственно, отображается сама на себя). Это условие сводится к замечательной гипотезе Хокинга о том, что черные дыры и их временные инверсии — белые дыры — с точки зрения физики неотличимы друг от друга! Аргументация Хокинга состояла в том, что в симметричной во времени физике состояние теплового равновесия тоже должно быть симметричным во времени. Я не хочу здесь пускаться в подробное обсуждение этой поразительной возможности. Идея Хокинга состояла в том, что квантово-механическое хокинговское излучение может рассматриваться как своего рода временная инверсия классического «поглощения» вещества черной дырой. Несмотря на всю изобретательность ее автора, эта гипотеза наталкивается на серьезные теоретические трудности и я лично не верю в ее работоспособность.

В любом случае, эта гипотеза плохо согласуется с теми идеями, которые я здесь выдвигаю. Я утверждал, что несмотря на то, что черные дыры должны существовать, существование белых дыр *запрещено гипотезой о вейлевской кривизне*! ГВК привносит в обсуждение элемент временной асимметрии, не учтенный Хокингом. Следует отметить, что поскольку черные дыры и их пространственно-временные сингулярности действительно занимают большое место в обсуждении того, что происходит внутри ящика Хокинга, то, следовательно, рассматриваемая проблема вне всякого сомнения тоже должна быть связана с неизвестными пока физическими законами, управляющими поведением такого рода сингулярностей. Хокинг считает, что эта неизвестная физика должна иметь вид *симметричной* во времени квантовой теории гравитации, а я считаю — что это должна быть *асимметричная* во времени ПКТГ! Я утверждаю, что одним из главных следствий ПКТГ должна быть ГВК (и, следовательно, второе начало термодинамики в известном нам виде), и поэтому необходимо попытаться установить, какие из ГВК вытекают следствия для рассматриваемой проблемы.

Посмотрим, как *включение* ГВК отразится на обсуждении потока «несжимаемой жидкости» в фазовом пространстве ℙ. Действие чернодырной сингулярности в пространстве-времени состоит в том, что она поглощает и разрушает все попадающее на нее вещество. Для нас же гораздо важнее,

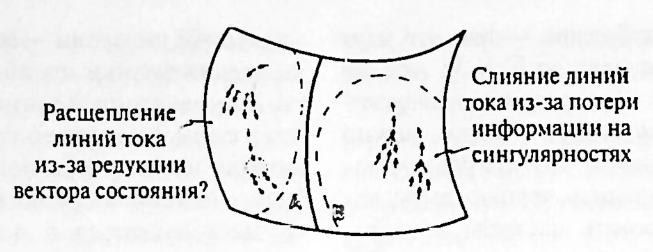


Рис. 8.6. В области $\mathbb{B}$ должно происходить слияние линий тока из-за потери информации на сингулярностях черных дыр. Компенсируется ли слияние порождением новых линий тока в результате **R**-процедуры (главным образом в области $\mathbb{A}$)?

что эта сингулярность *уничтожает информацию*! Следствием этого для фазового пространства $\mathbb{P}$ является слияние некоторых линий тока (рис. 8.6). Два состояния до этого различные могут превратиться в одно, как только различающая их информация окажется уничтоженной. При слиянии линий тока в фазовом пространстве $\mathbb{P}$ мы фактически имеем дело с *нарушением* теоремы Лиувилля. Наша жидкость больше не является несжимаемой, а *непрерывно уничтожается* в области $\mathbb{B}$!

Похоже, что теперь мы оказались в действительно трудном положении. Если «жидкость» постоянно уничтожается в области $\mathbb{B}$, то число линий тока из $\mathbb{A}$ в $\mathbb{B}$ должно *превышать* число линий тока из $\mathbb{B}$ в $\mathbb{A}$, откуда следует, что породить новую черную дыру легче, чем уничтожить уже имеющуюся! Все это действительно имело бы смысл, если бы не то обстоятельство, что теперь количество «жидкости», покидающее область $\mathbb{A}$, превышает количество «жидкости», которое возвращается в эту область. Черных дыр в области $\mathbb{A}$ нет, а существование белых дыр исключается ГВК — и поэтому теорема Лиувилля должна вне всякого сомнения абсолютно точно выполняться в области $\mathbb{A}$! Однако теперь, похоже, нам нужно каким-то образом «порождать жидкость» в области $\mathbb{A}$ для восполнения ее потери в области $\mathbb{B}$. Какой механизм может обеспечить увеличение числа линий тока? По-видимому, нам потребуется, чтобы в некоторых случаях одно и то же состояние могло приводить к более чем одному результату (т. е. допустить возможность бифуркации линий тока). Такого рода неопределенность эволюции физической системы в будущем «попахивает» квантовой теорией — ее **R**-частью. Возможно ли, чтобы **R** была в некотором смысле «оборотной стороной монеты» к ГВК? В то время, как ГВК обеспечивает слияние линий тока в области $\mathbb{B}$, квантово-механическая **R**-процедура приводит к бифуркациям линий тока. Я действительно утверждаю, что именно *объективный* квантово-механический процесс **R** редукции вектора состояния приводит к бифуркациям линий тока и таким образом в точности компенсирует их слияние, вызываемое ГВК (см. рис. 8.6)!

Для того, чтобы такое расщепление произошло, **R**-процедура должна быть, как мы уже видели, асимметричной во времени: вспомним описанный выше эксперимент с лампой, фотоэлементом и полупосеребренным зеркалом. В случае излучения лампой фотона возможны два (одинаково вероятных) результата этого процесса: либо фотон попадает на фотоэлемент и последний регистрирует его, либо фотон попадает на стену в точке A и фотоэлемент не срабатывает. В фазовом пространстве этого эксперимента мы имеем линию тока, представляющую излучение фотона, и эта линия тока расщепляется на две: одна часть представляет ситуацию, когда фотоэлемент срабатывает, а другая — когда он не срабатывает. Здесь мы, по-видимому, имеем дело с самой настоящей бифуркацией: одно допустимое состояние на входе и два возможных состояния на выходе. Второе входное состояние, которое следовало бы рассмотреть, — это испускание фотона из точки B на лабораторной стене, и в этом случае мы имели бы два состояния на входе и два на выходе. Однако только что упомянутое альтернативное состояние на входе исключается по причине его противоречия со вторым началом термодинамики — т. е. исходя из изложенной здесь концепции, и,

в конечном итоге, по причине противоречия с ГВК при отслеживании эволюции системы назад в прошлое.

Я должен еще раз отметить, что излагаемая мною здесь точка зрения на самом деле не является «традиционной» — хотя мне и не совсем понятно, как «традиционные» физики предлагают решать все поставленные здесь проблемы. (Я подозреваю, что немногие из них вообще серьезно над ними задумывались!) Разумеется, я слышал разные точки зрения. Например, время от времени некоторые физики выдвигали предположение о том, что хокинговское излучение никогда не приводит к *полному* исчезновению черной дыры, и что от нее всегда остается своего рода «ядрышко». (И, следовательно, согласно этой точке зрения стрелок из 𝔹 в 𝔸 нет!) На самом деле это почти никак не скажется на мои рассуждениях (и фактически даже усилит их). Можно, однако, избежать моих выводов, если постулировать, что общий объем фазового пространства ℙ на самом деле *бесконечен*, но это противоречило бы некоторым весьма фундаментальным представлениям об энтропии черных дыр и природе фазового пространства замкнутых (квантовых) систем, а другие технические приемы, позволяющие избежать моих выводов, о которых мне доводилось слышать, представляются еще менее удовлетворительными. Гораздо более серьезное возражение состоит в том, что построение ящика Хокинга требует слишком сильной идеализации, и что, предполагая возможность его создания, мы вынуждены преступать некоторые барьеры принципиального характера. Хотя я сам не до конца в этом уверен, но все же склоняюсь к тому, чтобы считать некоторую необходимую идеализацию вполне допустимой!

Наконец, есть один серьезный аспект, о котором я умолчал. Я начал обсуждение, предположив, что мы имеем дело с *классическим* фазовым пространством — а теорема Лиувилля относится к классической физике. Но затем пришлось рассмотреть квантово-механический феномен хокинговского излучения. (Кроме того, квантовая теория нужна для обеспечения *конечной размерности* и конечного объема ℙ.) Как мы видели в главе 6, квантовым аналогом фазового пространства является *гильбертово пространство*, и, поэтому, следовало бы, наверно, проводить все наши рассуждения в терминах гильбертова, а не фазового пространства. Для гильбертова пространства существует аналог теоремы Лиувилля, который следует из так называемого «*унитарного*» характера временной эволюции **U**. Не исключено, что все мои рассуждения можно сформулировать полностью в терминах гильбертового, а не классического фазового пространства, но мне трудно представить себе, каким образом в этом случае можно рассматривать классические явления, связанные с пространственно-временной геометрией черных дыр. Я считаю, что для правильной теории непригодно ни классическое фазовое пространство, ни гильбертово пространство, а потребуется какой-то новый, до сих пор еще не открытый тип математических пространств, занимающий промежуточное положение между двумя упомянутыми выше. Соответственно, мои рассуждения следует рассматривать только в эвристическом смысле, и они представляют собой скорее всего лишь общие *предположения*, а не окончательные выводы. Тем не менее, я действительно считаю свои рассуждения сильным доводом в пользу глубинной связи между ГВК и **R**, откуда вытекает, что **R**-процедура *действительно должна представлять собой эффект квантовой теории гравитации*.

Повторю свои выводы еще раз: я выдвигаю гипотезу, согласно которой квантово-механическая редукция вектора состояния действительно является оборотной стороной ГВК. В соответствии с этой гипотезой два важнейших следствия нашей искомой правильной квантовой теории гравитации (ПКТГ) — это ГВК и процедура **R**. ГВК приводит к слиянию линий тока в фазовом пространстве, в то время как процедура **R** приводит к *расщеплению* линий тока, в точности компенсирующему их слияние, вызванное ГВК. Оба процесса теснейшим образом связаны со вторым началом термодинамики.

Отметим, что слияние линий тока происходит только в области 𝔹, в то время как их расщепление может иметь место как внутри области 𝔸, также и внутри области 𝔹.

Вспомним, что 𝔸 представляет совокупность состояний, в которых черные дыры отсутствуют, и, следовательно, редукция вектора-состояния действительно возможна при отсутствии черных дыр. Ясно, что для выполнения R совсем необязательно иметь в лаборатории черную дыру (как в случае только что рассмотренного нами эксперимента с фотоном). Нас сейчас интересует лишь общий баланс между различными *возможными* событиями в той или иной ситуации. В рамках излагаемой концепции отсутствие детерминизма в квантовой теории должно всего лишь компенсироваться возможностью образования черных дыр на некотором этапе (и следующей отсюда возможностью уничтожения информации)!

Когда происходит редукция вектора-состояния?

Предположим, что мы признаем, исходя из вышеизложенных соображений, что редукция вектора-состояния может каким-то образом оказаться гравитационным феноменом. Можно ли сформулировать связь между R-процедурой и гравитацией более явным образом? Когда, согласно этой концепции, должен *фактически* иметь место коллапс вектора состояния?

Здесь следует прежде всего отметить, что даже в рамках более «традиционных» подходов к построению квантовой теории гравитации согласование принципов общей теории относительности с правилами квантовой механики наталкивается на определенные и весьма серьезные технические трудности. Эти правила (в первую очередь — интерпретация импульсов как дифференцирования по координатам в уравнении Шредингера — см. с. 254) плохо вписываются в представление об искривленной геометрии пространства-времени. Я лично считаю, что введение «значительной» пространственно-временной кривизны влечет неизбежное нарушение правил квантовой линейной суперпозиции. Именно в этом случае суперпозиция комплексных амплитуд в принципе допустимых альтернатив заменяется набором вероятностно-взвешенных реальных альтернатив, из которых одна *фактически* имеет место.

Что я понимаю здесь под «значительной» степенью кривизны? Я имею в виду достижение такой степени кривизны, при которой ее характерное значение становится сравнимым с *одногравитонным*[6] масштабом или превышает его. (Напомним, что, согласно правилам квантовой теории, электромагнитное поле «квантуется» на отдельные элементы, называемые «фотонами». При разложении поля на его частотные составляющие, компонента с частотой ν может входить в это разложение только в виде целого числа фотонов, каждый с энергией равной $h\nu$. Предполагается, что аналогичные правила должны быть также применимы и к гравитационному полю.) Один гравитон — это минимальная единица кривизны, допускаемая квантовой теорией. Идея состоит в том, что при достижении этого уровня обычные правила линейной суперпозиции, предписываемые процедурой U, должны претерпеть определенные изменения при их применении к гравитонам, и при этом возникает некая асимметричная во времени «нелинейная неустойчивость». Мы получаем вместо комплексных суперпозиций неограниченно долго сосуществующих «альтернативных возможностей» ситуацию, когда одна из «возможностей» начинает на этом этапе одерживать верх над другими и система «перескакивает» в то или иное из альтернативных состояний. Возможно, что выбор одного из альтернативных состояний происходит случайно, а быть может, в его основе лежат какие-то более глубокие законы. Однако теперь *реальность* обретает вид *одного* из альтернативных состояний. Процедура R осуществилась.

Отметим, что согласно этой гипотезе R-процедура осуществляется спонтанно, совершенно объективно и независимо от какого бы то ни было вмешательства человека. Идея состоит в том, что «одногравитонный уровень» должен находиться как раз между «квантовым уровнем» атомов, молекул и т. д., на котором хорошо действуют линейные правила (U) обычной квантовой механики, и «классическим уровнем» нашего повседневного опыта. Насколько «велик» одногравитонный уровень? Отметим, что дело тут на самом деле не в физическом размере, а скорее в распределении

массы и энергии. Как мы видели, эффекты квантовой интерференции могут возникать и на больших расстояниях при условии, что связанная с ними энергия мала. (Вспомним самоинтерференцию фотона, описанную на с. 228, и эксперименты типа ЭПР, проведенные Клаузером и Аспектом, с. 252.) Характерный масштаб массы в квантовой гравитации известен под названием планковской массы, приблизительно равной

$$m_{\text{Pl}} = 10^{-5} \text{ г.}$$

Она может показаться гораздо большей, чем хотелось бы, поскольку в простых наблюдениях мы видим, как гораздо менее массивные объекты, например, пылинки, ведут себя классическим образом. (Величина m_{Pl} немного меньше массы блохи.) Однако, я не думаю, что одногравитонный критерий применим столь грубым образом. Я постараюсь высказываться по-возможности яснее, но на момент написания этих строк вопрос о конкретном способе применения рассматриваемого критерия остается в значительной степени открытым.

Давайте рассмотрим сначала очень непосредственный способ наблюдения частицы — при помощи камеры Вильсона. В этом случае мы имеем камеру, заполненную паром, находящимся на грани конденсации в капельки воды. При попадании в такую камеру быстро движущейся частицы — например, частицы, возникшей в результате распада расположенного вне камеры радиоактивного атома, ее прохождение сквозь камеру вызывает ионизацию расположенных вблизи траектории пролета атомов (т. е. атомы становятся заряженными в результате отрыва от них электронов). Эти ионизированные атомы служат центрами конденсации капелек из водяного пара. Таким образом возникает трек, состоящий из капелек, которые могут непосредственно наблюдаться экспериментатором (рис. 8.7).

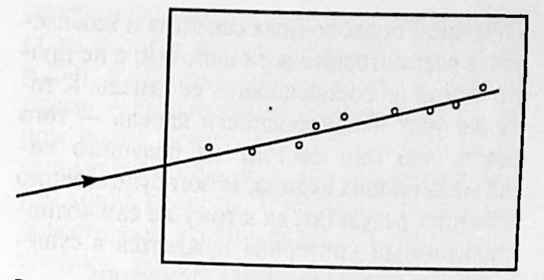


Рис. 8.7. Заряженная частица влетает в камеру Вильсона и вызывает конденсацию капелек на своем пути

Ну а как же все это описывается в квантовой механике? В момент распада радиоактивного атома он испускает частицу. Но у этой частицы существует множество различных направлений движения: каждое направление движения описывается своей амплитудой, причем все они сосуществуют одновременно в виде линейной квантовой суперпозиции. Совокупность всех этих наложенных друг на друга альтернатив образует исходящую из распавшегося атома сферическую волну — волновую функцию испущенной атомом частицы. При попадании любого из возможных треков частицы в камеру, он тут же оказывается ассоциированным с цепочкой ионизованных атомов, каждый из которых служит центром конденсации пара. Все эти различные возможные цепочки ионизованных атомов должны сосуществовать в виде линейной квантовой суперпозиции, так что мы имеем теперь линейную суперпозицию большого числа *различных* цепочек конденсирующихся капелек. На некотором этапе эта комплексная квантовая линейная суперпозиция превращается в действительную совокупность *фактических* альтернатив с вероятностными весами, равными, согласно R-процедуре, квадратам модулей амплитуд вероятностей. В реальном физическом мире реализуется только одна из этих альтернатив, и именно она наблюдается экспериментатором. В соответствии с излагаемой здесь точкой зрения эта стадия наступает, когда разность между гравитационными полями различных альтернативных вариантов достигает одногравитонного уровня.

Когда это происходит? Согласно очень грубым расчетам [7], если бы имелась только одна однородная шарообразная капля, то одногравитонный уровень достигался бы, когда ее масса вырастет до одной сотой от величины m_{Pl}, что составляет одну десятимиллионную грамма. В этом расчете много неопределенностей (включая трудности принципиального характера), да и величина полученной массы несколько великовата, однако результат не совсем уж бессмысленный. Остается надеяться на появление

в будущем более точных расчетов и возможность рассмотрения всей цепочки, а не просто одной из составляющих ее капель. К тому же учет неоднородности капель — того факта, что они состоят из большого числа мельчайших атомов, может существенно изменить результат, да к тому же сам «одногравитонный критерий» нуждается в существенном математическом уточнении.

В описанной выше ситуации рассматривалось возможное реальное наблюдение квантового процесса (распада радиоактивного атома), при котором квантовые эффекты оказываются усиленными настолько, что различные квантово-механические альтернативы приводят к различным и непосредственно наблюдаемым макроскопическим альтернативам. Я считаю, что **R**-процедура действительно может иметь место объективным образом даже в отсутствие столь *ярко выраженного* усиления. Предположим, что наша частица попала не в камеру Вильсона, а просто в большой ящик, заполненный газом (или жидкостью) с плотностью, обеспечивающей практически гарантированное столкновение частицы или иное ее воздействие на большое число атомов газа. Рассмотрим всего два варианта возможного поведения частицы, как составные части начальной линейной суперпозиции: частица может просто не попасть в ящик совсем или же она попадет в него по определенной траектории и окажется отраженной каким-либо атомом газа. Во втором случае соответствующий атом газа отскочит, двигаясь с очень большой скоростью так, как он никогда не повел бы себя, не столкнись он с частицей, затем столкнется с еще одним атомом и, в свою очередь, отрикошетит от него. После этого движение двух атомов будет отличаться от их движения в отсутствие столкновения с частицей, и мы будем иметь уже целый каскад движений атомов в газе, невозможный в отсутствие первоначального попадания частицы в ящик (рис. 8.8). Вскоре после этого порожденное частицей возмущение охватит практически все атомы газа.

Подумаем теперь, как эту ситуацию можно описать на языке квантовой механики. Вначале мы имеем лишь исходную частицу, и ее различные положения составляют комплексную линейную суперпозицию — волновую функцию частицы. Однако через какое-то время квантово-механическое описание должно уже охватывать все атомы газа. Рассмотрим комплексную суперпозицию двух возможных траекторий частицы, при движении по одной из которых частица попадает в ящик, а по другой — нет. Стандартная квантовая механика требует распространения этой суперпозиции на все атомы газа: мы должны рассмотреть суперпозицию двух состояний, таких, что положение атомов газа в одном состоянии оказываются смещенными относительно их положений в другом состоянии. Теперь рассмотрим *разность* гравитационных полей всех отдельных атомов. Хотя распределение газа (и гравитационное поле) *в целом* практически одинаково для обоих состояний, чью суперпозицию мы должны рассмотреть, если мы *вычтем* одно поле

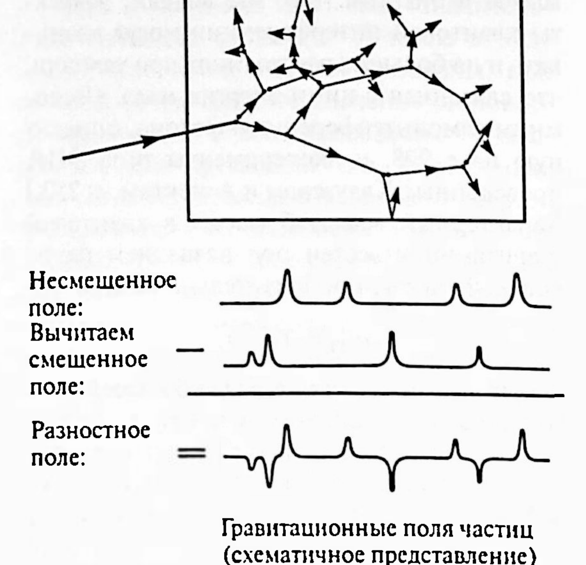


Рис. 8.8. Гравитационные поля частиц (условное изображение). При попадании частицы в ящик с газом через некоторое время практически все атомы газа оказываются охваченными порожденным частицей возмущением. Линейная квантовая суперпозиция частицы, попавшей в ящик, и частицы, не попавшей в ящик, подразумевает линейную суперпозицию двух различных пространственно-временных геометрий, описывающих гравитационные поля двух различных распределений частиц газа. В какой момент *различие* между этими геометриями достигает одногравитонного уровня?

из другого, то получим (сильно флуктуирующее) *разностное* поле, которое вполне может оказаться «значительным» в подразумеваемом здесь смысле — а именно это разностное поле вполне может превысить одногравитонный уровень. По достижении этого уровня немедленно же происходит редукция вектора состояния: в *реальном* состоянии частица *либо* попала в ящик, *либо* нет. Комплексная линейная суперпозиция сводится к статистически взвешенным альтернативам с осуществлением только *одной* из них.

В предыдущем примере я рассматривал камеру Вильсона в качестве способа квантово-механического наблюдения. Я считаю, что и другие виды таких наблюдений (фотопластинки, искровые камеры и т. д.) можно анализировать в рамках одногравитонного критерия, используя подход, примененный в описанном выше случае ящика с газом. Многое еще предстоит сделать, чтобы разобраться в подробностях применения этой процедуры.

Изложенные здесь соображения представляют собой всего лишь зачаток новой теории, которая, как мне кажется, является столь необходимой [8]. Для того, чтобы быть полностью удовлетворительной, любая схема должна, по-моему, включать в себя радикально обновленные представления о природе пространственно-временной геометрии, быть может, с применением нелокального описания [9]. Один из самых неоспоримых доводов в пользу этого следует из экспериментов ЭПР-типа (см. с. 247 и 252), в которых «наблюдение» (в данном случае — срабатывание фотоэлемента) в одном конце комнаты может вызвать мгновенную редукцию вектора-состояния в другом конце комнаты. Построение полностью объективной теории редукции вектора-состояния, не противоречащей духу теории относительности, представляет собой очень трудную и глубокую задачу, поскольку понятие «одновременности», будучи зависимым от движения некоторого наблюдателя, является чуждым теории относительности. Я убежден, что наше современное представление о физической реальности — особенно в том, что касается природы *времени* — нуждается в коренном пересмотре, пожалуй, даже в более радикальном, чем тот, который был вызван к жизни современной теорией относительности и квантовой механикой.

Вернемся к исходной проблеме. Какое все это имеет отношение к физическим законам, которые управляют действиями нашего мозга? Как это связано с нашими мыслями и чувствами? Для того, чтобы попытаться хоть как-то ответить на эти вопросы, придется сначала немного разобраться в устройстве нашего мозга. Потом я вернусь к проблеме, которую считаю фундаментальной: какого рода новые физические *действия* происходят, когда мы сознательно мыслим или воспринимаем что-либо?

Примечания

1. Вот самые распространенные корректировки этого типа: (i) замена уравнений Эйнштейна РИЧЧИ = ЭНЕРГИЯ (используя лагранжианы более высоких порядков); (ii) замена четырехмерного пространства-времени на пространство-время с бо́льшим числом измерений (как в случае так называемых «теорий Калуцы—Клейна»); (iii) введение «суперсимметрии» (идея, заимствованная из квантового поведения бозонов и фермионов, сведенного в единую схему, и примененная, не совсем последовательно, к пространственно-временны́м координатам); (iv) теория струн (очень популярная сейчас теория, в которой «мировые линии» заменяются на «истории струн» — обычно в сочетании с идеями (ii) и (iii)). Все эти предложения, несмотря на их популярность, следует рассматривать как заведомо ПРОБНЫЕ согласно терминологии главы 5.
2. Хотя процедуры квантования не всегда сохраняют симметрию классической теории (см. Трейман [1985]; Аштекар и др. [1989]), здесь требуется нарушение всех *четырех* симметрий, обычно обозначаемых как T, PT, CT и CPT. Это (особенно нарушение CPT симметрии) выходит за пределы возможностей обычных методов квантования.
3. Насколько я смог понять, именно такая точка зрения неявно содержится в выдвигаемых

сейчас Хокингом предложениях по квантово-гравитационному объяснению рассматриваемых проблем (Хокинг [1987, 1988]). Гипотеза Хартли и Хокинга [1983] о квантово-гравитационной природе начального состояния, возможно, относится к тем гипотезам, что *могут подвести* теоретическую базу под начальное условия типа ВЕЙЛЬ = 0, но эти идеи пока что лишены *чрезвычайно важного* (по моему мнению) компонента, каким является асимметрия во времени.

4. Это станет несколько более понятным, если использовать операцию скалярного произведения $\langle\psi|\chi\rangle$, упомянутую в примечании 6 к главе 6. В случае описания вперед по времени вероятность p рассчитывается как:

$$p = \left|\langle\chi'|\psi'\rangle\right|^2 = \left|\langle\psi'|\chi'\rangle\right|^2.$$

Тождественность двух выражений следует из $\langle\psi'|\chi'\rangle = \langle\psi|\chi\rangle$, а это, в сущности, и подразумевается под «унитарной эволюцией».

5. Возможно, некоторым читателям сложно понять, что имеется в виду под вероятностью прошлого события при условии, что имело место определенное событие в будущем. Однако это совсем не сложно. Вообразите себе всю историю нашей вселенной, отображенной в пространстве-времени. Чтобы найти вероятность события p при условии, что произошло событие q, мысленно рассмотрим все случаи, когда имело место событие q, и сосчитаем, в какой доле этих случаев имело место также и событие p. Это и есть требуемая вероятность. При этом не важно, относится ли q к событиям, которые обычно происходят после события p, или до него.

6. Следует допустить, что это как раз и есть так называемые *продольные гравитоны* — «виртуальные» гравитоны, из которых состоит статическое гравитационное поле. К сожалению, четкое и «инвариантное» математическое определение таких объектов связано с определенными теоретическими трудностями.

7. Мои собственные первые грубые расчеты этой величины были очень существенно улучшены Абхеем Аштекаром, и здесь я привожу значение, определенное Аштекаром (см. Пенроуз [1987a]). Аштекар, однако, специально отметил, что многие из предположений довольно произвольны, и поэтому следует относиться к полученному значению массы весьма осторожно.

8. Время от времени в литературе появляются и другие попытки построения объективной теории редукции векторов состояний. Среди наиболее существенных следует отметить работы Каройхази [1974], Каройхази, Френкеля и Лукача [1986], Комара [1969], Перла [1985, 1989], Гирарди, Римини и Вебера [1986].

9. На протяжении нескольких лет я тоже пытался разрабатывать нелокальную теорию пространства-времени, побуждаемый к этому главным образом стимулами иного рода, исходящими из так называемой «теории твисторов» (см. Пенроуз, Риндлер [1986], Хаггетт, Тод [1985], Уорд, Уэллс [1990]). Однако этой теории в лучшем случае недостает ряда существенных ингредиентов, и обсуждение ее здесь представляется неуместным.

Глава 9

Реальный мозг и модели мозга

Как же устроен мозг?

У нас в голове находится великолепное устройство, которое управляет нашими действиями и каким-то образом дает нам представление об окружающем мире. Правда, как однажды отметил Алан Тьюринг [1], внешне оно больше всего напоминает миску холодной овсянки! Трудно представить, как столь заурядного вида объект умудряется совершать чудеса, на которые, как мы знаем, он способен. Однако при ближайшем рассмотрении оказывается, что он имеет гораздо более сложное строение и замысловатую организацию (рис. 9.1).

Бóльшая покрытая извилинами (и более всего похожая на овсянку) часть мозга, расположенная сверху, называется собственно *головным* (*или большим*) *мозгом*. Он четко делится посередине на правое и левое *полушария* и, более условно, в передне-заднем направлении — на лобную долю и три остальные: височную, теменную и затылочную. Еще дальше и несколько книзу расположен небольшой и округлый *мозжечок*, чем-то похожий на пару клубков шерсти. Глубоко внутри мозга, как бы укрытый им, находится целый ряд любопытных и сложных на вид структур: варолиев мост и продолговатый мозг, которые вместе с ретикулярной формацией — областью, к которой мы обратимся позднее — составляют ствол мозга, а также таламус, гипоталамус, гиппокамп, мозолистое тело и еще много других, странных как по виду, так и по названиям, частей.

Большой мозг — предмет особой гордости человека и не только потому, что он является самой большой частью человеческого мозга, но и потому, что пропорция между этой частью и мозгом в целом у *человека* больше, чем у животных. (Мозжечок человека тоже превосходит размерами таковой у большинства других животных.) Головной мозг и мозжечок имеют сравнительно тонкий наружный слой *серого вещества*, под которым расположено значительно большее по массе *белое вещество*. Эти области серого вещества называют, соответственно, корой головного мозга и корой мозжечка. Считается, что в сером веществе происходят различные вычислительные действия, а белое вещество, состоящее из длинных нервных волокон, отвечает за передачу сигналов из одной части мозга в другую.

Каждой из различных областей коры головного мозга присущи свои специфические функции. Зрительная кора расположена в затылочной доле, прямо в задней части мозга, и занимается восприятием и распознаванием зрительных образов. Забавно, что природа именно там решила разместить интерпретатор визуальной информации, получаемой зрительными органами, которые (по крайней мере, у человека) находятся прямо спереди! Но природа вытворяет и куда более странные вещи. Так, за *левую половину* человеческого тела практически полностью отвечает *правое* полушарие, тогда как за правую — почти исключительно левое, поэтому чуть ли не все нервы, идущие в головной мозг или выходящие из него, по необходи-

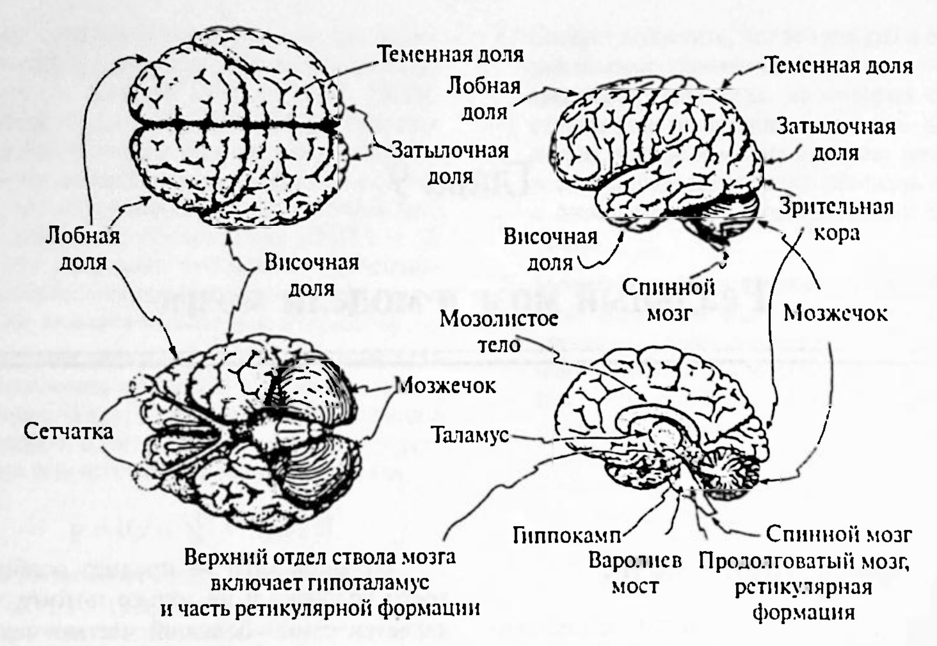


Рис. 9.1. Мозг человека: вид сверху, сбоку, снизу и в разрезе

мости должны перекрещиваться! При этом в случае зрительной коры правая ее часть связана не с левым глазом, а с *левой частью поля зрения обоих глаз*. Аналогично, левая часть зрительной коры связана с правой частью поля зрения обоих глаз.

Это означает, что нервы от правой части сетчатки каждого из глаз должны идти к правой половине зрительной коры (вспомните, что изображение на сетчатке перевернуто по отношению к источнику), а нервы от левой части сетчатки — к левой половине коры (рис. 9.2). Таким образом в левой и правой частях зрительной коры формируется четкое отображение правой и левой областей поля зрения, соответственно.

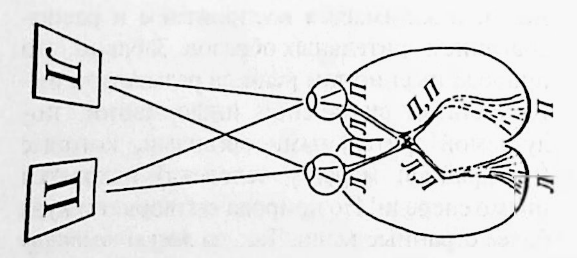


Рис. 9.2. Левая сторона поля зрения обоих глаз отображается на правой половине зрительной коры, а правая, соответственно, на левой (вид снизу; обратите внимание, что предметы на сетчатке отображаются в перевернутом виде)

Сигналы от ушей приходят на противоположные части мозга столь же замысловатым образом. Правая слуховая кора (часть правой височной доли) обрабатывает в основном звуки, поступающие слева, а левая слуховая кора — звуки, поступающие справа. Обоняние кажется здесь исключением из общего правила. Правая часть обонятельной коры, которая расположена в передней части большого мозга (в передней доле — что уже само по себе является исключением для сенсорной области), отвечает в основном за правую ноздрю, а левая часть — за левую ноздрю.

Осязание связано с областью затылочной доли мозга, которая носит название *соматосенсорной* коры. Эта область находится как раз за условной границей, разделяющей лобную и теменную доли. Между различными частями поверхности тела и отдельными участками соматосенсорной коры существует довольно своеобразное соответствие. Иногда оно изображается графически в виде так называемого «соматосенсорного гомункулуса» — искаженной человеческой фигуры, изображаемой лежащей вдоль соматосенсорной коры, как это показано на рис. 9.3.

Правая часть соматосенсорной коры принимает осязательные сигналы, идущие

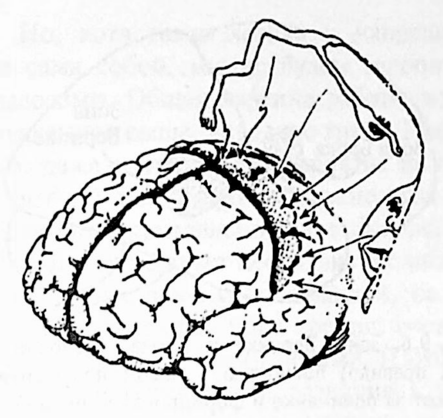


Рис. 9.3. «Соматосенсорный гомункулус» наглядно иллюстрирует участки коры головного мозга, расположенные сразу за линией, разделяющей лобную и теменную доли и непосредственно связанные с частями тела, откуда поступает осязательная информация

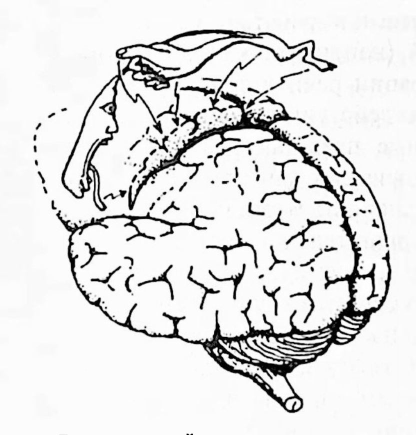


Рис. 9.4. «Двигательный гомункулус» наглядно изображает участки коры головного мозга, примыкающие к линии раздела между лобной и теменной долями, которые непосредственно приводят в движение различные части тела

от левой стороны тела, а левая — с правой. В лобной доле непосредственно *перед* границей с теменной долей находится участок коры, известный как *двигательная кора*. Он приводит в движение различные части нашего тела. И опять мы встречаемся с точно определенным соответствием между мышцами нашего тела и зонами этого участка мозга. Как и в случае с осязанием, эти связи можно графически изобразить в виде «двигательного гомункулуса» (рис. 9.4). И снова правая часть двигательной коры отвечает за движение левой стороны тела, а левая — правой.



Рис. 9.5. Функции большого мозга (грубая схема). Сенсорная информация извне поступает в первичную область восприятия, последовательно обрабатывается до мельчайших деталей во вторичной и третичной сенсорных областях, затем передается в третичную двигательную область, и, в конце концов, в первичных двигательных областях преобразуется в точные инструкции к действию

Все упомянутые выше зоны коры головного мозга (зрительная, слуховая, обонятельная, осязательная и двигательная) называются *первичными*, поскольку именно они непосредственно осуществляют прием поступающих в мозг и передачу исходящих из него сигналов. Рядом с ними расположены *вторичные* зоны, предназначенные для более тонкой и сложной обработки сенсорной информации (рис. 9.5).

Сенсорная информация, полученная зрительной, слуховой или соматосенсорной зоной коры головного мозга, обрабатывается соответствующими вторичными областями, после чего вторичная двигательная область вырабатывает план движения, который переводится первичной двигательной областью на язык прямых команд, непосредственно адресованных мышцам. (Мы не будем касаться обонятельного участка коры, поскольку он функционирует иным и малоизученным пока образом.) Остальные участки коры головного мозга относятся к разряду *третичных* (или *ассоциативных*). В этих областях в основном и выполняется наиболее сложная и характеризуемая высокой степенью абстрагирования часть умственной

деятельности. Именно здесь при определенном участии периферической нервной системы собирается воедино и подвергается всестороннему анализу информация, поступающая от различных сенсорных участков; здесь происходит запоминание, складываются картины внешнего мира, намечаются и оцениваются планы действий, распознается и генерируется речь.

Речь представляет для нас особый интерес, поскольку ее обычно относят к разряду способностей, присущих исключительно человеческому интеллекту. Интересно, что (по крайней мере у подавляющего большинства правшей и большей части левшей) *речевые* центры находятся в основном в *левой* половине мозга. К важным участкам относятся *зона Брока*, расположенная в задней нижней части лобной доли, и *зона Вернике*, которая располагается внутри и вокруг верхней задней части височной доли (рис. 9.6). Зона Брока отвечает за построение предложений, а зона Вернике — за понимание языка. Повреждение зоны Брока приводит к нарушению речи, но не ее пониманию, тогда как при повреждении зоны Вернике речь остается беглой, но, в основном, бессмысленной. Пучок нервных волокон, который связывает между собой две эти области, называется *дуговидным пучком*. При его повреждении ни речь, ни ее понимание не нарушаются, но мысль не может быть выражена словами.

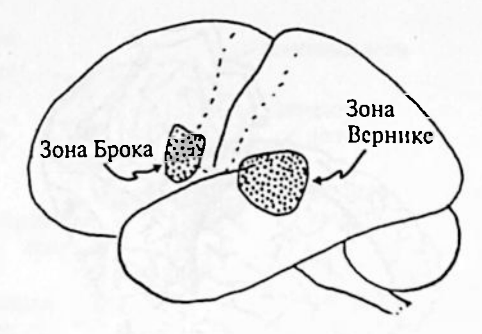


Рис. 9.6. Зоны Вернике и Брока, расположенные (как правило) полностью в левом полушарии, отвечают за понимание и формирование речи соответственно

Мы теперь можем составить очень приблизительную картину того, что делает головной мозг. *Входные данные* для мозга представляют собой зрительные, слуховые, осязательные и прочие сигналы, которые сначала регистрируются в первичных областях (главным образом) *задних* долей (теменной, височной и затылочной). *Выходные сигналы* мозга, приводящие к различным движениям тела, вырабатываются в основном *лобными* долями мозга. А где-то между ними происходит обработка информации и принятие решений. В общем, можно сказать, что активность мозга, начавшись в первичных областях задних долей, перемещается затем во вторичные области, где входные данные анализируются, и, далее, в третичные области задних долей, где информация становится полностью осмысленной (как, например, в случае с пониманием речи в зоне Вернике). *Дуговидный пучок* — упомянутый выше пучок нервных волокон, но теперь уже с обеих сторон мозга, — переносит эту информацию в лобную долю, где ее третичными областями вырабатывается общий план действий (например, как это происходит при генерации речи в зоне Брока). Эти общие планы действий преобразуются в более конкретные представления о движениях тела во вторичных двигательных областях, откуда активность мозга перемещается в первичную двигательную кору, которая, в конце концов, посылает соответствующие сигналы различным группам мышц тела (и часто нескольким одновременно).

Создается впечатление, что перед нами предстает картина превосходного вычислительного устройства. Сторонники сильного ИИ (см. главу 1 и далее) рассматривают мозг как великолепный образец алгоритмического компьютера — по сути, машины Тьюринга — в котором есть входные данные (как на ленте слева от машины Тьюринга) и выходные данные (как на ленте справа от машины Тьюринга) и который способен выполнять всевозможные нетривиальные вычисления на промежуточных этапах. Конечно, активность мозга может не прекращаться и в отсутствие внешних раздражителей. Это происходит в тех случаях, когда человек думает, занимается вычислениями или предается воспоминаниям. Приверженцы сильного ИИ отнесли бы это на счет продолжающейся алгоритмической деятельности и предположили бы, что явление «осознания» возникает как раз в те моменты, когда подобная деятельность достигает определенного уровня сложности.

Но, хотя такая логика и напрашивается сама собой, мы не будем торопиться с выводами. Общая картина работы мозга, приведенная выше, довольно груба. Прежде всего, даже зрительное восприятие не происходит по такой простой схеме, как это было мной представлено ранее. В коре, по-видимому, существует несколько различных (хотя и менее значимых) областей, на которые отображаются поля зрения, очевидно, с какими-то другими целями. (Похоже, именно они отвечают за различия в том, как мы *осознаем* увиденное.) Скорее всего, по коре разбросаны также и другие дополнительные сенсорные и двигательные области (например, движение глаз может быть вызвано сигналами из определенных точек *задних* долей).

В своем описании мозга я затронул только его кору и ни разу не коснулся вопроса о назначении прочих частей. Какую роль выполняет, например, *мозжечок*? Ясно, что он отвечает за координацию и контроль движений тела, его равновесие, своевременность и точность действий. Представьте себе артистичность танцора, отточенность движений профессионального игрока в теннис, мгновенную реакцию гонщика, уверенные движения рук музыканта или художника; подумайте о грациозных прыжках газели или крадущейся кошке. Без мозжечка подобная точность движений была бы невозможна, они стали бы неуверенными и неуклюжими. По-видимому, в процессе приобретения новых навыков, будь то ходьба или вождение машины, сначала человеку приходится детально обдумывать каждое свое действие, и за это отвечает кора головного мозга, но когда достигнут определенный уровень мастерства и действия начинают выполняться «автоматически», управление ими передается мозжечку. Более того, хорошо известно, что как только профессионал *задумывается о* своих действиях, он на время теряет легкость их координации. *Думание*, по-видимому, сопровождается переходом контроля к коре головного мозга и, хотя при этом, как следствие, появляется гибкость действий, «мозжечковая» плавность и точность движений на время утрачивается. Такое описание, без сомнения, является чересчур упрощенным, но тем не менее позволяет нам в общих чертах понять функцию мозжечка[1)].

При описании функций головного мозга до сих пор вообще не упоминалось о других частях мозга. Например, *гиппокамп* играет важнейшую роль в формировании долговременной (постоянной) памяти, хотя сама память располагается где-то в коре головного мозга, возможно, во многих местах одновременно. Мозг способен также сохранять образы различными способами с помощью кратковременной памяти в течение нескольких минут или даже часов (просто, что называется, «держа их в голове»). Но для того, чтобы человек мог вспомнить эти образы после того, как его внимание с них переключилось, необходимо сохранить их в долговременной памяти, и здесь уже не обойтись без гиппокампа. (Повреждение этого участка мозга приводит к ужасному состоянию, когда человек не способен запомнить ничего нового и все сразу забывается, как только его внимание переключается на другой объект.) *Мозолистое тело* — это область, ответственная за связь между двумя полушариями мозга. (Далее мы увидим, к каким поразительным явлениям приводит рассечение мозолистого тела.) *Гипоталамус* представляет собой эмоциональный центр, в котором гнездятся удовольствие, ненависть, страх, отчаяние, голод, и который служит посредником между эмоциями и их ментальными и физическими проявлениями. Между гипоталамусом и различными частями мозга идет постоянный обмен сигналами. *Таламус* функционирует как важный обрабатывающий центр и переключающий узел, который передает значительную часть импульсов, поступающих извне, в кору головного мозга. *Ретикулярная формация* отвечает за общее состояние готовности мозга и его отдельных частей к осознанному восприятию. Все эти и многие другие жизненно важные части мозга соединены многочисленными нервами.

Вышеприведенное описание дает только общее представление о некоторых наиболее значимых частях мозга. Мне кажется целесообразным в завершение этого раздела

[1)] Интересно, что для мозжечка не характерно «перекрестное» поведение коры головного мозга: правая половина мозжечка управляет, в основном, правой стороной тела, а левая — левой.

привести некоторые сведения о строении мозга в целом. Его различные части группируются в три отдела, которые, если двигаться от позвоночника, называются по порядку *задним* (*rhombencephalon*), *средним* (*mesencephalon*) и *передним* (*prosencephalon*) мозгом. На ранних стадиях развития эмбриона эти отделы, в том же порядке, видны как три вздутия на конце позвоночного столба. Самое дальнее — развивающееся в передний мозг — имеет два выроста в виде пузырей, по одному с каждой стороны, которые становятся большими полушариями головного мозга. Полностью развитый передний мозг включает в себя многие важные части всего мозга — не только большой головной, но и мозолистое тело, таламус, гипоталамус, гиппокамп и многие другие. Мозжечок является частью заднего мозга. Ретикулярная формация расположена частью в среднем мозге, а частью в заднем. Передний мозг является «новейшим» отделом с точки зрения эволюционного развития, а задний — наиболее «древним».

Я надеюсь, что это краткое описание, во многом неточное, даст читателю некоторое представление о том, на что похож мозг человека и как он функционирует. До сих пор я лишь вскользь упомянул то, что служит центральной темой нашей дискуссии — *сознание*. Теперь перейдем к этому вопросу вплотную.

Где обитает сознание?

Существует множество различных точек зрения на соотношение между состоянием мозга и феноменом сознания. Насколько очевидна важность этого явления, настолько же велико и расхождение во взглядах на него. Однако ясно, что не все части мозга в равной степени участвуют в формировании сознания. Например, как следует из вышесказанного, мозжечок по роду своей деятельности гораздо ближе к «автоматическому устройству», чем кора головного мозга. Действия, контролируемые мозжечком, происходят как будто сами собой и не требуют «обдумывания». Когда мы сознательно решаем пройти от одного места до другого, то вряд ли имеем перед собой тщательно разработанный план мышечных сокращений, который был бы необходим для управляемого движения. То же самое можно сказать и о бессознательных рефлекторных действиях, как, например, отдергивание руки от горячей печи, которое может быть опосредовано не головным мозгом, а верхней частью спинного мозга. Таким образом, напрашивается вывод о том, что феномен сознания, вероятнее всего, связан с активностью головного мозга, а не мозжечка или спинного мозга.

С другой стороны, совершенно не очевидно, что активность коры головного мозга всегда определяет осознанность наших действий. Например, как я уже указывал, в норме при ходьбе человек *не* контролирует детальные движения конечностей и работу мышц — управление этими действиями осуществляет, в основном, мозжечок (с помощью других частей головного мозга и спинного мозга), — однако первичные двигательные области головного мозга *тоже* вовлекаются в этот процесс. Более того, то же можно сказать и о первичных сенсорных областях: мы можем совершенно не осознавать меняющееся при ходьбе давление на подошвы ног, тем не менее соответствующие участки соматосенсорной коры постоянно активируются.

Уайлдер Пенфилд, выдающийся американо-канадский нейрохирург (среди заслуг которого — составление в 1940-х и 1950-х годах детальных карт двигательных и сенсорных областей мозга человека), считал, что сознание не связано просто с активностью коры головного мозга. На основании опыта проведения многочисленных операций на мозге пациентов, находившихся в сознании, он предположил, что область, которую он называл *верхней частью ствола мозга*, включающая, в основном, таламус и средний мозг (см. Пенфилд, Джаспер [1947]), хотя он имел в виду главным образом ретикулярную формацию, в некотором смысле может быть названа «центром сознания». Верхняя часть ствола мозга связана с корой головного мозга, и, согласно Пенфилду, «акт осознания» или «осознанное действие» происходит каждый раз, когда эта область ствола мозга непосредственно обменивается сигналами с определенным участком коры, отвечающим именно

за те чувства, мысли, воспоминания или действия, которые в данный момент осознанно воспринимаются или совершаются. Он указывал, что можно, например, стимулировать определенный участок двигательной коры мозга, который отвечает за движение правой руки (и правая рука на самом деле будет двигаться), но это не вызовет у подопытного *желания* двигать правой рукой. (Более того: он может даже постараться остановить ее движение левой рукой — совсем как доктор Стрэнджлав из популярного фильма![2]) Пенфилд предполагал, что *желание* совершить действие связано скорее с таламусом, нежели с корой головного мозга. Согласно его представлениям сознание — это проявление активности верхней части ствола мозга, однако, поскольку должно еще быть что-то, *что* осознается, то эта активность не ограничивается стволом мозга, но включает в себя еще и те участки коры, с которыми у верхней части ствола мозга в этот момент существует активная связь и которые представляют собой субъект (чувственное восприятие или воспоминание) или объект (волевое действие) сознания.

Другие нейрофизиологи тоже высказывали предположение о том, что ретикулярную формацию можно было бы назвать «местонахождением» сознания, если таковое на самом деле существует. Ведь, как бы там ни было, эта область отвечает за пребывание мозга в активном состоянии. Ее повреждение приводит к потере сознания. Всегда, когда мозг находится в бодрствующем сознательном состоянии, активна и ретикулярная формация, и наоборот. На самом деле существует явная связь между активностью ретикулярной формации и тем состоянием человека, которое мы традиционно называем «сознательным». Однако ситуация осложняется тем, что во сне, когда мы на самом деле «сознаем», что мы спим, активные в норме участки ретикулярной формации активности не проявляют. И еще один факт мешает ученым признать за ретикулярной формацией столь почетный статус: с точки зрения эволюции, эта часть мозга является очень *древней*. Если все, что нужно для обладания сознанием — активность ретикулярной формации, то им должны быть наделены лягушки, ящерицы и даже треска!

Лично я не расцениваю последний довод как достаточно весомый. Разве у нас есть неоспоримые свидетельства того, что ящерицы и треска *не* обладают неким зачаточным сознанием? Какое право мы имеем утверждать, как это некоторые делают, что человеческие существа — единственные обитатели нашей планеты, наделенные свыше настоящим «сознанием»? Неужели на Земле мы единственные, кому дозволено «осознавать»? Позвольте усомниться в этом. Конечно, лягушки, ящерицы и уж тем более треска не вызывают у меня ощущения, что «кто-то в них» взирает на меня, когда я рассматриваю эти создания, но я очень явственно ощущаю присутствие «сознания», когда смотрю в глаза кошке, собаке или, особенно, когда на меня смотрят обезьяны или мартышки в зоопарке. Я не требую, ни чтобы они чувствовали то же, что и я, ни даже какой-либо сложности испытываемых ими чувств. Им совершенно не обязательно «сознавать себя» в каком-то строгом смысле этого слова (хотя наличие некоторого элемента самосознания у них я не исключаю[3]). Достаточно будет, чтобы они просто *чувствовали*! Что касается состояния сна, то я бы признал, что определенная форма сознания при этом присутствует, хотя, по всей видимости, на довольно низком уровне. Если за функционирование сознания каким-то образом отвечают только участки ретикулярной формации, то они должны сохранять активность (хотя бы невысокую) и во время сна.

Другая точка зрения (О'Кифи [1985]) состоит в том, что сознание в большей мере связано с функционированием *гиппокампа*. Как я уже отмечал, гиппокамп определяет способность к долговременному запоминанию. Принимая в качестве гипотезы, что

---

[2] Речь идет о фильме *Доктор Стрэнджлав*, в котором Питер Сэллер играет нацистского врача — доктора Стрэнджлава, — эмигрировавшего в США и вынужденного все время останавливать левой рукой свою правую руку, которая самовольно вскидывается в нацистском приветствии. — *Прим. ред.*

[3] О том, что, по крайней мере, шимпанзе обладают самосознанием, с убедительностью говорят результаты экспериментов, в ходе которых шимпанзе разрешалось играть с зеркалами (см. Окли [1985], главы 4 и 5).

постоянная память связана с сознанием, мы должны рассматривать гиппокамп как главное действующее лицо в феномене осознанного восприятия.

Есть и другое мнение, согласно которому сознание является результатом деятельности самой коры головного мозга. Раз уж большой головной мозг служит предметом особой гордости человека (хотя у дельфинов он никак не меньше!), и умственная деятельность, понимаемая как интеллект, связана как раз с этой частью мозга, то именно в ней и должна обитать душа человека! Таким, по-видимому, мог бы быть вывод, например, сторонников сильного ИИ. Если «осознание» — не более, чем следствие *сложности* алгоритма — или, возможно, его «глубины» или некой «степени изощренности», — тогда, в соответствии с представлениями сильного ИИ, сложные алгоритмы, выполняемые корой головного мозга, дали бы ей преимущественное право претендовать на способность к проявлению сознания.

Многие философы и психологи склонны считать *язык* непременным атрибутом человеческого сознания. Соответственно, именно способность изъясняться при помощи слов позволяет достичь той тонкости мышления, которая служит отличительной чертой человека и выражением самой его сути. Именно язык, в соответствии с этой точкой зрения, отличает нас от других животных и дает нам возможность лишать их свободы и вести на бойню, как только в этом возникает потребность. Именно язык позволяет нам философствовать и описывать наши ощущения, так что мы можем убедить остальных, что *мы* осознаем окружающий мир и самих себя. С этой точки зрения владение языком является необходимым и достаточным условием наличия сознания.

А теперь мы должны вспомнить о том, что языковые центры находятся (у большинства людей) в *левой* половине мозга (зоны Брока и Вернике). Из вышеизложенной точки зрения должно было бы следовать, что сознание — это что-то, что связанное только с левой половиной коры головного мозга! И таково, на самом деле, мнение целого ряда нейрофизиологов (в частности, Джон Экклз [1973]), которое я, как сторонний наблюдатель, считаю весьма странным по причинам, изложенным ниже.

Эксперименты при разделенных больших полушариях мозга

В связи со сказанным выше я должен упомянуть целый ряд замечательных наблюдений над людьми (и животными) при полном рассечении у них мозолистого тела, которое делало взаимодействие левого и правого полушарий головного мозга невозможным. У людей [2] операция по рассечению мозолистого тела применялась как эффективное средство лечения в случаях особо тяжелых форм эпилепсии. Роджер Сперри с сотрудниками подвергал таких пациентов, спустя некоторое время после операции, многочисленным психологическим тестам. При этом в левом и правом полях зрения испытуемых помещались никак не связанные друг с другом предметы, так что левое полушарие получало информацию только о том, что располагалось с правой стороны, а правое полушарие — с левой. Если справа предъявлялось изображение карандаша, а слева — чашки, то тестируемый произносил: «Это *карандаш*», поскольку именно карандаш, а не чашку, воспринимала та половина мозга, которая явно отвечает за речевые способности. Однако левой рукой испытуемый выбирал блюдце, а не лист бумаги, считая его ассоциативно более подходящим к чашке. Левая рука находилась «в подчинении» у правого полушария, которое, хотя и не могло оперировать словами, все же было способно производить определенные, довольно сложные и типичные для человека действия. Было высказанно предположение о том, что за «геометрическое мышление» (особенно пространственное воображение) и музыкальное восприятие ответственно, в основном, *правое* полушарие, а за речевые и аналитические способности — левое. Правое полушарие мозга может понимать общеупотребительные существительные и элементарные предложения, а также выполнять простейшие арифметические действия.

Самое поразительное, что при разделении полушарий они ведут себя как две практически независимые индивидуальности, с каждой из которых экспериментатор может общаться по отдельности, хотя общение с правым полушарием носит более примитивный характер и значительно затруднено по сравнению с левым из-за отсутствия речевых способностей. Каждая половина головного мозга может поддерживать связь с другой половиной косвенным путем, например, наблюдая за движениями руки, контролируемыми другой стороной, или слыша звуковые «подсказки» (такие, как стук блюдца). Но в хорошо контролируемых лабораторных условиях даже эта примитивная связь может быть устранена. Однако от одной половины к другой все же могут передаваться неясные эмоциональные ощущения, предположительно потому, что нерассеченные структуры мозга, такие как гипоталамус, по-прежнему связаны с обоими полушариями.

Возникает искушение задать вопрос: неужели перед нами — два различных индивидуума, обладающих сознанием и пребывающих в одном теле? Этот вопрос вызвал бурную полемику. Одни без сомнений отвечали на этот вопрос утвердительно, другие считали, что ни одна из сторон не должна рассматриваться как полноценная личность. Некоторые утверждали, что общность эмоциональных ощущений может служить доказательством существования только одной личности. Еще одна точка зрения состоит в том, что сознательного индивидуума представляет только *левое* полушарие, а правое — просто автомат. Этой точки зрения придерживаются те, кто считает речевые способности обязательной составляющей сознания. Само собой, только левое полушарие может убедительно заявить «Да!» в ответ на вопрос: «Обладаешь ли ты сознанием?». Правому полушарию, подобно кошке, собаке или шимпанзе, может быть трудно даже понять отдельные слова этого вопроса, не говоря уже о том, чтобы правильно ответить на него.

И все же пока вопрос остается открытым. В недавних экспериментах, проведенных Дональдом Вильсоном и его коллегами (Вильсон и др. [1977], Газзанига и др. [1977]), при наблюдениях за пациентом с разделенным мозгом (назовем его "P. S."), были получены весьма интересные результаты. После операции по разделению полушарий только левое полушарие обладало речью, но понимали речь *оба* полушария, а позднее правое полушарие научилось и воспроизводить речь! Несомненно, что оба полушария были наделены сознанием. Более того, это были два *отдельных* сознания, поскольку их желания и пристрастия были совершенно различны. Например, левое полушарие выражало желание стать чертежником, а правое — гонщиком!

Лично я не верю в справедливость широко распространенного убеждения в том, что обычный человеческий язык необходим для мышления или сознания. (В следующей главе я приведу некоторые доводы в пользу своей точки зрения.) Поэтому я отношусь к тем, кто верит, что, в принципе, обе половины мозга после разделения обладают сознанием независимо друг от друга. Пример с P. S. может служить весомым подтверждением тому, что, по крайней мере в этом частном случае, это так и есть. По-моему мнению, единственное действительное различие между P. S. и всеми другими случаями заключается в том, что сознание его правого полушария на самом деле смогло убедить окружающих в своем существовании!

Если мы допускаем, что P. S. действительно имеет два независимых разума, то возникает довольно пикантная ситуация. Есть все основания полагать, что *до* операции разделения полушарий у каждого пациента было только одно сознание. Однако после операции их уже два! В некотором смысле, изначально единственное сознание *раздвоилось*. Мы можем в связи с этим вспомнить гипотетического путешественника из главы 1 (с. 58), который воспользовался телепортационной машиной и в какой-то момент (неумышленно) был поставлен перед фактом, что будто бы «настоящее» его «я» благополучно прибыло на Венеру. В этом случае раздвоение сознания приводит к кажущемуся парадоксу. Ведь мы можем задать резонный вопрос: «А какой, собственно, маршрут выбрал поток его сознания „на самом деле"?» Если бы *вы* были этим путешественником, то какое бы из двух

сознаний вы, в конце концов, назвали бы «собой»? Устройство для телепортации относится к области научной фантастики, однако в случае с P. S. мы имеем в чем-то аналогичную ситуацию и притом *совершенно реальную*! Какое из сознаний P. S. было бы правомерно «отождествить» с P. S. до операции? Нет сомнений, что многие философы сочли бы этот вопрос бессмысленным, ибо его решение при помощи операционалистских методов кажется невозможным. Каждое полушарие сохраняет память о «дооперационных» временах, и, естественно, каждое будет идентифицировать себя с той — еще целостной — личностью. Но все же подобная ситуация, примечательная в качестве своего рода головоломки и способная поставить в тупик, сама по себе еще не является парадоксальной.

Эта головоломка еще усложнится, если предположить, что в дальнейшем оба сознания можно было бы каким-то образом опять свести воедино. Повторное соединение разрезанных нервных волокон мозолистого тела на сегодняшнем этапе развития медицины исключается, но можно представить себе некий способ разделения полушарий, более мягкий, чем реальное разрезание нервных волокон. Например, нервные волокна могли бы быть временно заморожены или парализованы при помощи лекарственных средств. Пока я не слышал о подобных опытах, но думаю, что появление технических возможностей для их осуществления — это вопрос обозримого будущего. Тогда можно допустить, что после приведения мозолистого тела в работоспособное состояние, мы вновь получим *одно* сознание! Представьте, что это сознание ваше. Как бы вы себя чувствовали после того, как в течение какого-то времени были двумя независимыми личностями с отдельными «я»?

«Зрение вслепую»

Эксперименты по разделению полушарий мозга, помимо прочего, ясно показали, что наличие *единственного* «места для сознания» вовсе не обязательно. Но были проведены и другие опыты, результаты которых дают основание полагать, что некоторые участки коры головного мозга в большей степени связаны с сознанием, нежели прочие. Среди подобных опытов — изучение явления *слепоты*. Повреждение тех или иных участков зрительной коры может привести к слепоте в соответствующем секторе поля зрения. Человек не видит предмет, помещенный в этот сектор — у него появляется частичная слепота, связанная с этой конкретной зоной его поля зрения.

Однако, кое-какие любопытные изыскания (см. Вайскранц [1987]) позволяют говорить о том, что дела здесь обстоят совсем не так просто, как кажется. У пациента, называемого здесь и далее "D. B.", необходимо было удалить часть зрительной коры головного мозга, и после операции у него наступила частичная слепота в описанном выше смысле. Однако, когда что-либо (как правило, изображение крестика, кружочка или наклонного отрезка прямой) помещали в «слепую зону» и просили D. B. *угадать*, что это такое, он обнаружил, что может делать это с практически стопроцентной точностью! Эта способность к «угадыванию» оказалась неожиданной и для самого D. B., который при этом продолжал утверждать, что в этой зоне он вообще ничего не видит[4)].

Изображения, формируемые на сетчатке, в свою очередь тоже обрабатываются не только зрительной корой, но и *другими* участками мозга, при этом один из наиболее загадочных из них находится в нижней части височной доли. Вполне возможно, что D. B. строил свои «догадки» на основе информации, полученной как раз этим участком нижней части височной доли. При активации этих областей не возникало никаких *осознанных* ощущений, однако информация в них, бесспорно, содержалась, проявляя себя только в точности «догадок» D. B. На самом деле, после соответствующей тренировки D. B. научился до некоторой степени осознавать информацию, относящуюся к этим областям мозга.

[4)] Своего рода дополнительным к «зрению вслепую» может служить состояние, известное как «отрицание слепоты», при котором совершенно слепой человек настаивает на том, что он хорошо видит, и которое, по-видимому, связано с *визуальным* осознанием информации об окружении, полученной при помощи *других* органов чувств! См. Черчланд [1984], с. 143.

Все это, по-видимому, указывает на то, что отдельные зоны коры головного мозга (как, например, зрительная кора) имеют большее отношение к сознательному восприятию, чем другие, но некоторые из этих менее важных зон, очевидно, могут быть путем тренировок открыты для непосредственного доступа сознания.

Обработка информации в зрительной коре

Именно в *зрительной коре* процессы обработки информации изучены гораздо лучше, чем в других частях мозга. Для их описания был предложен целый ряд разнообразных моделей [3]. На самом деле, *до того*, как визуальная информация попадает в зрительную кору, ее частичная обработка проходит еще в сетчатке. (Вообще говоря, сетчатка считается частью мозга!) Одни из первых экспериментов по исследованию процессов обработки информации в зрительной коре были проведены Давидом Хьюбелом и Торстеном Визелем и в 1981 году принесли им Нобелевскую премию. В ходе этих экспериментов удалось показать, что определенные клетки зрительной коры кошки воспринимают в поле зрения линии, имеющие вполне *определенный угол наклона*. При этом соседние клетки были восприимчивы к линиям, расположенным под несколько иным углом. Часто не имело значения, что именно характеризуется таким углом наклона. Это могла быть граница между темной и светлой областью или просто темная черта на светлом фоне. Изучаемые клетки оказались способны абстрагироваться от конкретной природы объекта, имеющего свойство «угол наклона». Другие клетки были чувствительны к определенным цветам или к различиям между изображениями, регистрируемыми каждым глазом, что позволяет воспринимать объемные изображения. Продвигаясь далее от первичных областей восприятия, мы обнаруживаем клетки, которые чувствительны ко все более тонким аспектам восприятия того, что мы видим. Например, при взгляде на рис. 9.7 мы различаем очертания белого треугольника, однако линии, образующие сам треугольник, большей частью не изображены,

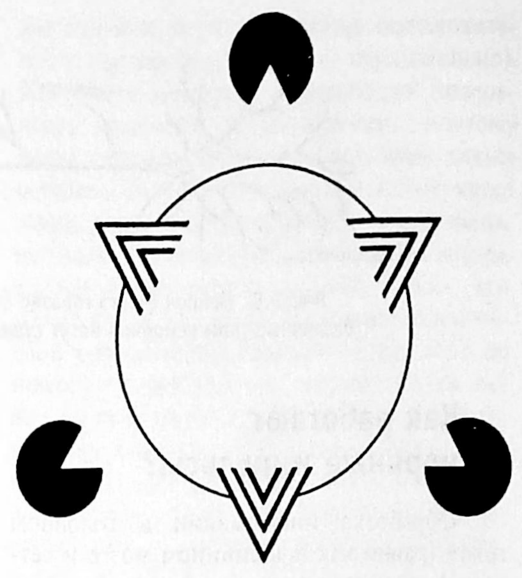


Рис. 9.7. Видите ли вы белый треугольник, лежащий поверх другого треугольника и прикрепленный к нему кольцом? Стороны этого белого треугольника нигде не нарисованы до конца, однако, в мозге есть клетки, ответственные за восприятие этих невидимых линий

но *домыслены*. Клетки, способные фиксировать эти «подразумеваемые» линии, действительно были обнаружены в зрительной коре (той, что называется вторичной зрительной корой)!

В начале 1970-х годов в литературе [4] появились заявления об открытии в зрительной коре мозга мартышек клеток, которые активируются только тогда, когда на сетчатку проецируется изображение *лица*. На основании этой информации была сформулирована «гипотеза бабушкиного нейрона», согласно которой в мозге человека должны существовать определенные клетки, реагирующие только в тех случаях, когда в комнату входит его/ее бабушка! Недавние исследования показали, что есть клетки, реагирующие на определенные слова. Может быть, это шаг на пути к доказательству справедливости гипотезы бабушкиного нейрона?

Ясно, что нам предстоит еще очень много узнать о деталях процессов обработки информации в мозге. До сих пор очень мало известно о функционировании высших отделов мозга. Мы пока оставим эти вопросы и обратимся к самим клеткам мозга, которые позволяют ему осуществлять эту удивительную деятельность.

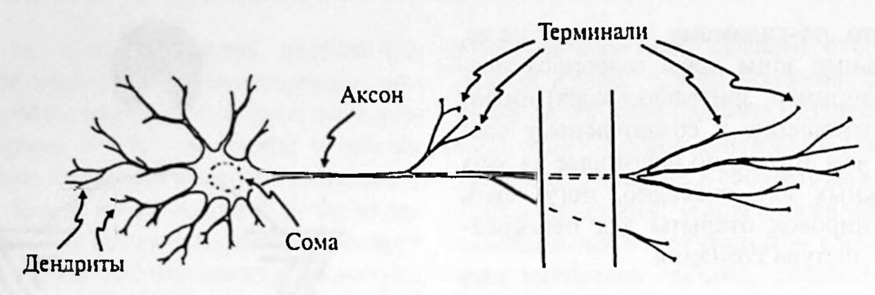


Рис. 9.8. Нейрон (часто гораздо более удлиненный, чем на рисунке). Различные типы нейронов могут существенно отличаться по внешнему виду

Как работают нервные импульсы?

Обработка информации в головном мозге (равно как и в спинном мозге и сетчатке) осуществляется уникальными по своему разнообразию клетками, которые называются *нейронами* [5]. Попробуем разобраться, как же устроен нейрон. Я схематично изобразил его на рис. 9.8. Его утолщенная центральная часть, немного похожая на звезду и часто имеющая форму редиски, называется *телом* (*сомой*) *нейрона* и содержит в себе клеточное ядро. С одной стороны от тела нейрона отходит сильно вытянутое нервное волокно, называемое *аксоном*. Аксон иногда достигает действительно огромной длины (у человека — часто до нескольких сантиметров), если учесть, что речь идет всего лишь об одной микроскопической клетке.

Аксон служит «проводом», по которому передается *исходящий* из клетки нервный сигнал. От аксона в стороны могут отходить более мелкие ветви и, кроме того, аксон может несколько раз разветвляться. На концах каждого из этих нервных волокон находятся *нервные окончания* (*терминали*). По другую сторону сомы, а часто и отходя от нее во всех направлениях, располагаются короткие сильно ветвящиеся отростки — *дендриты*, по которым в клетку поступают *входные* данные. (Иногда и на концах дендритов встречаются терминали, образующие так называемые *дендро-дендритные* синапсы между дендритами. В дальнейшем я не буду их учитывать, поскольку связанное с ними усложнение общей картины несущественно.)

Клетка как целое отделена от окружения клеточной мембраной, которая охватывает сому, аксон, нервные окончания, дендриты и все остальное. Для того, чтобы сигналы передавались от одного нейрона к другому, надо каким-то образом обеспечить им возможность «перехода через барьер» между нейронами. Это достигается с помощью межклеточного соединения, называемого *синапсом*, в котором терминаль одного нейрона соединена с какой-либо точкой на соме или на одном из дендритов другого нейрона (рис. 9.9). На самом деле, между терминалью одного нейрона и сомой или дендритом другого остается очень узкий зазор, который называется *синаптической щелью* (рис. 9.10). При передаче от одного нейрона к другому сигнал должен преодолеть этот зазор.

В какой форме сигналы передаются по нервным волокнам и через синаптические щели? Что заставляет следующий нейрон передавать сигнал дальше? Для непосвященного, вроде меня, механизмы, которые используются здесь природой, кажутся удивительными и совершенно зачаровывающими!

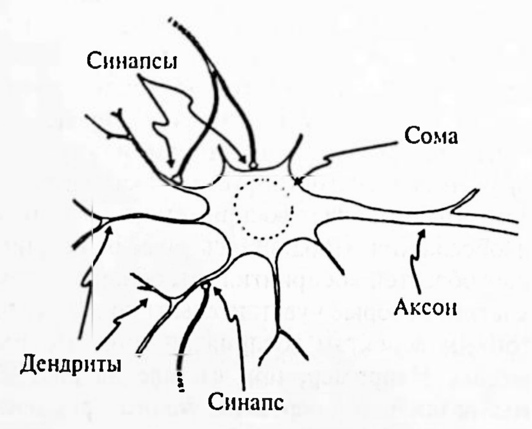


Рис. 9.9. Синапсы обеспечивают контакт одного нейрона с другим

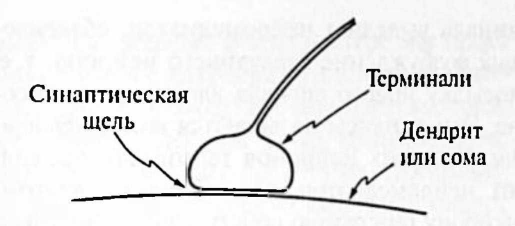


Рис. 9.10. Схема строения химического синапса. Через синаптическую щель сигнал передается с помощью нейромедиатора

Можно было бы думать, что эти сигналы распространяются точно так же, как электрический ток по проводам, но в действительности все гораздо сложнее.

Нервное волокно представляет собой цилиндрическую трубку, заполненную раствором обычной соли (хлорида натрия), смешанной с хлоридом калия (с преобладанием последнего), так что внутри трубки находится смесь из ионов натрия, калия и хлора (рис. 9.11). Снаружи волокна находятся те же ионы, но в других соотношениях: ионов натрия больше, чем ионов калия. В состоянии покоя содержимое трубки имеет суммарный отрицательный заряд (т. е. ионов хлора там больше, чем ионов калия и натрия вместе; напомним, что ионы калия и натрия заряжены положительно, тогда как ионы хлора — отрицательно). Клеточная мембрана, образующая поверхность цилиндра, имеет «утечки», поэтому ионы перемещаются через мембрану таким образом, чтобы нейтрализовать избыточный заряд. Компенсацию утечек и поддержание избыточного отрицательного заряда внутри трубки осуществляет «ионный насос», который очень медленно откачивает ионы натрия через мембрану наружу. Отчасти это же помогает поддерживать избыток ионов калия по сравнению с ионами натрия во внутреннем растворе. Существует также ионный насос, который (более медленно) переносит ионы калия из наружной среды внутрь трубки (что, правда, не способствует поддержанию разности зарядов).

Сигнал, распространяющийся по нервному волокну, представляет собой область с *обратным* распределением зарядов (т. е. положительный заряд внутри и отрицательный снаружи), которая перемещается вдоль волокна (рис. 9.12). Вообразите, что вы находитесь на нервном волокне как раз перед такой областью с обратным распределением зарядов. По мере того, как эта область

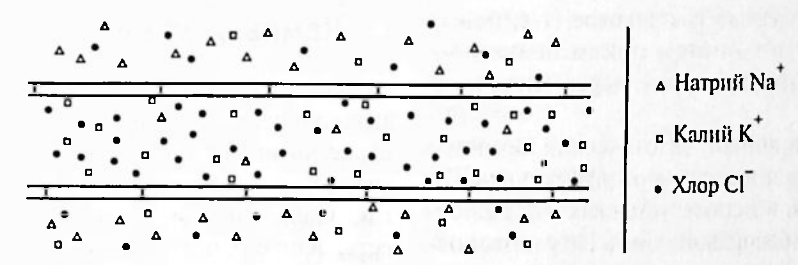


Рис. 9.11. Схематическое изображение нервного волокна. В состоянии покоя внутри волокна ионов хлора больше, чем ионов калия и натрия, что обеспечивает отрицательный суммарный заряд; снаружи ситуация противоположная, и, соответственно, имеется положительный заряд. Калиево-натриевый баланс внутри трубки отличается от баланса снаружи: внутри больше ионов калия, а снаружи — натрия

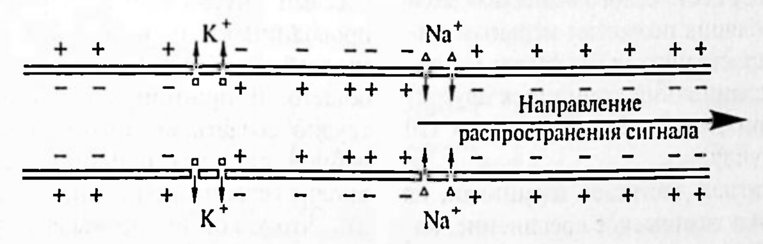


Рис. 9.12. Нервный импульс — это область с обратным (по отношению к состоянию покоя) распределением заряда, перемещающаяся вдоль волокна. При ее приближении открываются натриевые каналы, пропускающие поток ионов натрия внутрь; сразу после ее прохождения открываются калиевые каналы, обеспечивающие отток ионов калия наружу. Работа ионных насосов восстанавливает исходное состояние

приближается, электрическое поле открывает в мембране маленькие «дверцы», называемые *натриевыми каналами*. Это позволяет ионам натрия перемещаться с наружной стороны мембраны обратно внутрь трубки (в результате совместного действия электрических сил и давления, обусловленного разностью концентраций, т. е. «осмоса»). Это приводит к тому, что заряд снаружи становится отрицательным, а внутри — положительным. Когда это происходит, мы знаем, что область обратного распределения заряда, которая и является сигналом, достигла нас. При этом позади нее открываются крошечные «дверцы» другого типа (*калиевые каналы*), которые выпускают ионы калия наружу, тем самым восстанавливая избыточный отрицательный заряд внутри. Теперь сигнал прошел! Наконец, когда сигнал уже достаточно удалился, медленно, но верно работающие ионные насосы постепенно выкачивают ионы натрия из трубки наружу, закачивая внутрь ионы калия. Таким образом волокно возвращается в состояние покоя и готово к передаче очередного сигнала.

Обратите внимание, что сигнал представляет собой просто область обратного распределения заряда, движущуюся вдоль волокна. *Вещество* как таковое (т. е. ионы) перемещается при этом совсем немного — только внутрь и наружу через клеточную мембрану!

Этот странный, экзотической механизм действует на поверку очень эффективно. Он универсален и используется как у позвоночных, так и у беспозвоночных. Но у позвоночных он был усовершенствован за счет изоляции нервных волокон при помощи беловатого жироподобного вещества, называемого *миелином*. (Именно миелиновым покрытием объясняется цвет «белого вещества» мозга.) Такая изоляция позволяет нервным импульсам распространяться без потерь (от одной «ретрансляционной станции» к другой) и с очень приличной скоростью — до 120 метров в секунду.

Когда сигнал достигает терминали, из нее выделяется химическое соединение, называемое нейромедиатором. Это соединение пересекает синаптическую щель и достигает другого нейрона — поверхности дендрита или сомы. При этом у одних нейронов терминаль выделяет нейромедиатор, *облегчающий* возбуждение следующего нейрона, т. е. посылку нового сигнала вдоль своего аксона. Эти синапсы называются *возбуждающими*. У других нейронов терминали выделяют нейромедиатор, *затрудняющий* другому нейрону генерацию собственного импульса, и поэтому называются *тормозящими*. На каждом нейроне действие активных в данный момент возбуждающих синапсов суммируется, из результата вычитается суммарное действие тормозящих синапсов, и если полученная разность превышает определенное критическое значение, то нейрон действительно возбуждается. (Возбуждающие синапсы создают положительную разность потенциалов между внутренней и наружной сторонами мембраны следующего нейрона, а тормозящие — отрицательную. Эти разности потенциалов складываются. Нейрон возбудится только в том случае, если результирующая разность потенциалов на мембране в начале его аксона достигнет определенной критической величины, при которой ионы калия не успевают выходить наружу достаточно быстро, чтобы восстановить равновесие.)

Компьютерные модели

Важным свойством нервной системы является то, что сигналы, используемые для передачи информации, относятся (большей частью) к классу явлений «все или ничего». Сила сигнала не изменяется: он или есть, или его нет. Это придает деятельности нервной системы некоторое сходство с работой цифрового компьютера. На самом деле, между работой огромного количества взаимосвязанных нейронов и процессами внутри компьютера со всеми его проводниками и логическими элементами (подробнее об этом чуть позже) есть много общего. В принципе, было бы не так уж трудно создать компьютерную модель подобной системы нейронов. Но возникает вполне естественный вопрос: не означает ли это, что какой бы ни была детальная схема соединений нейронов в мозге, всегда можно построить его компьютерную модель?

Чтобы сделать это сравнение более наглядным, я должен объяснить, что такое

логический элемент. В компьютере мы также сталкиваемся с ситуацией типа «все или ничего»: либо в проводнике есть импульс тока, либо его нет, причем когда импульс *есть*, его величина всегда одна и та же. Поскольку все в компьютере строго синхронизовано, то *отсутствие* импульса было бы определенным сигналом, который может быть «замечен» компьютером. Вообще говоря, когда мы пользуемся термином «логический элемент», мы неявно подразумеваем, что наличие или отсутствие импульса обозначает «истину» или «ложь», соответственно. Конечно же, к реальной истине или лжи это никакого отношения не имеет и используется только как общепринятая терминология. Мы будем также обозначать «истину» (наличие импульса) цифрой «**1**» и «ложь» (отсутствие импульса) цифрой «**0**». Помимо этого, как и в главе 4, мы будем обозначать знаком «&» логическое «и» (которое является «утверждением» об «истинности» обоих аргументов, т. е. принимает значение **1** тогда и только тогда, когда оба они равны **1**); «∨» — логическое «или» (которое «означает», что либо один из аргументов, либо оба они «истинны», т. е. выражение становится равным **0** тогда и только тогда, когда оба аргумента имеют значение **0**); знаком «⇒» — «следует» (т. е. **A** ⇒ **B** означает утверждение «если истинно **A**, то истинно **B**»; «⇔» — «тогда и только тогда» (выражение истинно, если оба аргумента «истинны» или же оба «ложны» одновременно); и использовать знак «∼» для логического «не» (выражение «истинно», если аргумент «ложен», и «ложно», если аргумент «истинен»). Результаты применения различных логических операций можно описать при помощи так называемых «таблиц истинности»:

$$\mathbf{A\&B}: \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \qquad \mathbf{A \vee B}: \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

$$\mathbf{A \Rightarrow B}: \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \mathbf{A \Leftrightarrow B}: \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

в каждой из которых **A** обозначает строки (т. е. $\mathbf{A} = \mathbf{0}$ дает первую строку, а $\mathbf{A} = \mathbf{1}$ — вторую), а **B** — столбцы. Например, если $\mathbf{A} = \mathbf{0}$ и $\mathbf{B} = \mathbf{1}$, что во всех таблицах отвечает правому верхнему углу, то выражение **A** ⇒ **B** согласно *третьей* таблице примет значение **1**. (Соответствующий словесный пример из области *традиционной* логики: утверждение «если я сплю, то я счастлив», очевидно, остается истинным в частном случае, когда я бодрствую и счастлив.) И, наконец, действие логического элемента «не» может быть записано просто как:

$$\sim \mathbf{0} = \mathbf{1} \quad \text{и} \quad \sim \mathbf{1} = \mathbf{0}.$$

Это — основные типы логических элементов. Есть еще ряд других, но все они могут быть построены из только что описанных [6].

Итак, можем ли мы, в принципе, построить компьютер, используя соединенные между собой *нейроны*? Я собираюсь показать, что это возможно даже при самых примитивных представлениях о функциях нейрона. Посмотрим, как можно было бы, в принципе, построить логические элементы на основе соединенных между собой нейронов. Нам потребуется новый способ записи цифр, поскольку в *отсутствие* сигнала ничего не происходит. Будем считать (совершенно произвольно), что двойной импульс обозначает **1** (или «истину»), а одиночный — **0** (или «ложь»). Примем также упрощенную схему, в которой нейрон возбуждается только при получении *двух* возбуждающих импульсов (т. е. двойного импульса) одновременно. Тогда нетрудно сконструировать элемент «и» (т. е. «&»). Как показано на рис. 9.13, для этого достаточно, чтобы с выходным нейроном образовывали входные синапсы два нервных окончания. (Тогда, если по обоим окончаниям приходят двойные импульсы, то и первый, и второй импульс превысят заданный двухимпульсный порог срабатывания; а если хотя бы на один входной синапс приходит одиночный импульс, то превысит порог лишь одна

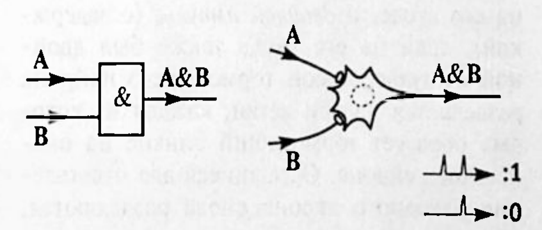


Рис. 9.13. Логический элемент «и». В «нейронной модели» (справа) нейрон возбуждается только в том случае, когда на его вход поступают одновременно два импульса

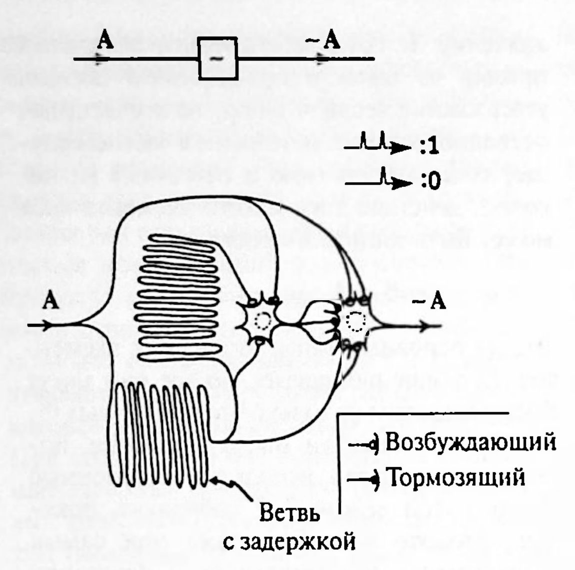


Рис. 9.14. Логический элемент «не». В «нейронной модели», как и ранее, для срабатывания нейрона требуется одновременное воздействие двух (по крайней мере) одиночных импульсов

пара возбуждающих импульсов. Я предполагаю, что все импульсы хорошо согласованы по времени, и что в случае двойного импульса, для определенности, синхронизация осуществляется по *первой паре* импульсов.)

Конструкция элемента «не» (т. е. «~») значительно сложнее. Один из способов его построения приведен на рис. 9.14. Входной сигнал поступает по аксону, разделяющемуся на две ветви. Одна из ветвей имеет увеличенную длину, такую, чтобы сигнал при движении по ней запаздывал ровно на время, равное промежутку между импульсами в паре. Затем обе ветви снова разделяются, и одно из ответвлений каждой ветви отходит к тормозящему нейрону, причем аксон от ветви с задержкой предварительно разделяется снова, образуя прямую ветвь и ветвь с задержкой. На выходе тормозящего нейрона не будет *ничего* при одиночном импульсе на его входе, и *двойной импульс* (с задержкой), если на его входе также был двойной импульс. Аксон тормозящего нейрона разделяется на три ветви, каждая из которых образует тормозящий синапс на оконечном нейроне. Оставшиеся два ответвления исходного аксона снова разделяются, так что к конечному нейрону подходят уже четыре терминали, образующие возбуждающие синапсы. При желании читатель может проверить, что выходной сигнал этого конечного нейрона соответствует сигналу элемента «не» (т. е. пара импульсов, если на входе был одиночный, и наоборот). (Такая конструкция кажется абсурдно усложненной, но это наилучшее из того, что пришло мне в голову!) В качестве развлечения читатель может составить подобные «нейронные» схемы и для остальных описанных выше логических элементов.

Естественно, эти конкретные примеры не могут служить серьезными моделями того, что происходит в мозге на самом деле. С их помощью я только старался показать, что описанная выше модель возбуждения нейрона по сути логически эквивалентна конструкции электронного компьютера. Легко видеть, что с помощью компьютера можно воспроизвести любую модель соединения нейронов между собой. В то же время, подробно рассмотренные выше конструкции указывают на то, что и, наоборот, системы нейронов могут быть моделями компьютера и, следовательно, *могут* действовать как (универсальная) машина Тьюринга. Хотя при обсуждении машин Тьюринга во второй главе мы не использовали понятие логических элементов [7] и, в действительности, для построения модели машины Тьюринга в общем случае помимо логических элементов нам понадобилось бы еще многое другое, в этом нет ничего принципиально нового, если только мы допускаем возможность *аппроксимации* используемой в машине Тьюринга *бесконечной ленты* огромным, но конечным множеством нейронов. А это уже, как кажется, подводит нас к выводу о том, что мозг по своей сути эквивалентен компьютеру!

Но прежде, чем делать такие поспешные выводы, нам следует рассмотреть некоторые различия между деятельностью мозга и работой современных компьютеров, которые могут оказаться достаточно важными. Во-первых, я слишком упростил описание возбуждения нейрона, отнеся его к явлениям типа «все или ничего». Это справедливо для одиночного импульса, распространяющегося по аксону. На самом деле, когда нейрон возбуждается, он генерирует целую последовательность импульсов, быстро следующих друг за другом. Даже в состоянии покоя нейрон генерирует импульсы,

но с гораздо меньшей частотой. Именно многократное увеличение *частоты* импульсов характеризует переход нейрона в возбужденное состояние. Кроме того, есть еще и вероятностный аспект срабатывания нейрона. Один и тот же стимул может приводит к различным результатам. Более того, в мозге нет точной синхронизации с помощью постоянной тактовой частоты, которая необходима для работы современных компьютеров. Кроме того, следует отметить, что максимальная частота срабатывания нейрона, составляющая около 1000 импульсов в секунду, гораздо меньше, чем у современных электронных устройств, у которых она более чем в 1 млн раз выше. К тому же, по сравнению с очень высокой точностью соединений в электронном компьютере, действительные соединения между нейронами кажутся в большой степени случайными и избыточными — правда, сегодня мы знаем, что в мозге (при рождении) эти соединения установлены с гораздо большей точностью, чем считалось полвека назад.

Может показаться, что большая часть из сказанного выше характеризует мозг с невыгодной стороны по сравнению с компьютером. Но есть и другие факторы, говорящие в пользу мозга. У логических элементов может быть лишь очень ограниченное количество входов и выходов (скажем, три-четыре, не больше), тогда как нейроны могут иметь гигантское число синапсов. (Предельным случаем можно считать нейроны мозжечка, известные как клетки Пуркинье, у которых количество возбуждающих синапсов достигает 80 000.) Помимо этого, общее число нейронов в мозге также превышает максимальное количество транзисторов, входящих в состав самой большой в мире вычислительной машины — примерно 10^{11} в мозге и «всего лишь» 10^9 у компьютера. Однако последнее число в будущем, скорее всего, возрастет [8]. Более того, большое число клеток мозга в значительной степени обусловлено огромным количеством мелких *клеток-зерен* в мозжечке, которых насчитывается около тридцати миллиардов (3×10^{10}). Если считать, что осознанным восприятием, в отличие от современных компьютеров, мы обладаем просто благодаря большому числу нейронов, то нам придется найти какое-то дополнительное объяснение тому, что деятельность *мозжечка* полностью *бессознательна* и в то же время сознание может быть связано с *головным мозгом*, в котором нейронов всего в два раза больше (около 7×10^{10}) при значительно меньшей плотности.

Пластичность мозга

Между деятельностью мозга и работой компьютера существуют и другие различия, на мой взгляд даже более важные, чем до сих пор упоминавшиеся, и связанные с явлением, которое называется *пластичностью мозга*. В действительности, неправомерно рассматривать мозг как *фиксированную* совокупность связанных друг с другом нейронов. Взаимосвязи нейронов на самом деле не постоянны, как это было бы в рассмотренной выше компьютерной модели, но все время меняются. Это не значит, что изменяются положения аксонов или дендритов. Многие из их сложных взаимосвязей в общих чертах формируются еще при рождении. Я имею в виду синаптические контакты, которые в действительности и обеспечивают связь между нейронами. На дендритах они часто формируются на небольших выростах, называемых *шипиками*, к которым подходят терминали других нейронов (рис. 9.15). Здесь «контакт» означает *не* соприкосновение, а узкий зазор (синаптическую щель) заданной ширины — около одной сорокатысячной доли миллиметра. При определенных условиях шипики дендритов могут

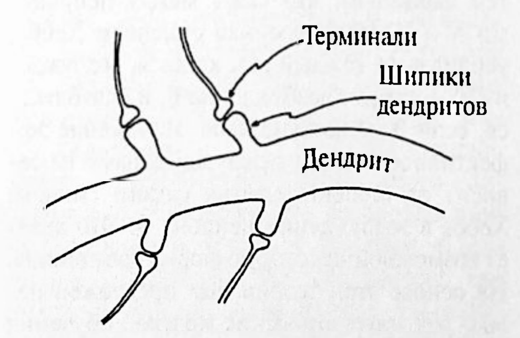


Рис. 9.15. Синаптические контакты, образуемый шипиками дендритов. Эффективность такого соединения легко изменяется при росте или уменьшении шипика

исчезать, тем самым нарушая контакт, или вырастать (могут образовываться и новые) и формировать новую связь. Таким образом, если мы представим себе, что совокупность соединенных друг с другом нейронов в мозгу действительно образует компьютер, то это компьютер, способный непрерывно изменяться!

Согласно одной из ведущих теорий долговременная память обусловлена именно такими изменениями синаптических контактов. Именно они обеспечивают возможность сохранения необходимой информации. Если это так, то пластичность предстает перед нами уже не просто как несущественное усложнение деятельности мозга, но как ее важнейшее свойство.

Каков механизм этих непрекращающихся изменений? Как быстро они могут происходить? Однозначный ответ на второй вопросов вряд ли существует, хотя представители по крайней мере одной из научных школ утверждают, что такие изменения могут происходить за несколько секунд. Этого можно было ожидать, если такие изменения ответственны за долговременное запоминание, поскольку оно происходит за характерное время около одной секунды (Кандел [1976]). Это имело бы для нас весьма существенное значение в дальнейшем. Я вернусь к этому важному вопросу в следующей главе.

А что же можно сказать о механизмах пластичности мозга? Согласно оригинальной теории, предложенной в 1954 году Дональдом Хеббом, существуют определенные синапсы (впоследствии получившие название «синапсов Хебба»), обладающие тем свойством, что связь между нейронами А и В, обусловленная синапсом Хебба, усиливается каждый раз, когда за возбуждением А следует возбуждение В, и ослабляется, если В не возбуждается. Изменение эффективности связи между нейронами не зависит от степени участия самого синапса Хебба в возбуждении нейрона В. Это делает возможной некоторую форму «обучения». На основе этой теории был предложен целый ряд математических моделей обучения и решения задач. Они получили название *нейронных сетей*. По-видимому, нейронные сети действительно способны к какому-то элементарному обучению, но им пока еще далеко до реальных моделей мозга. В любом случае, механизмы, управляющие изменениями синаптических контактов, скорее всего более сложны, чем рассмотренные выше. Очевидно, что необходимы дальнейшие исследования.

С пластичностью связан и другой аспект выделения нейромедиаторов терминалями. Иногда нейромедиаторы выделяются вовсе не в синаптические щели, а в окружающую межклеточную жидкость, возможно, для воздействия на другие, расположенные на большом удалении нейроны. По-видимому, многие нейрохимические вещества выделяются подобным образом. Существуют различные теории памяти, в которых используются разнообразные сочетания таких веществ, участвующих в процессе запоминания. Конечно, состояние мозга зависит от наличия в нем химических соединений (например, гормонов), выделяемых различными его частями. Проблемы нейрохимии в целом весьма сложны, и пока непонятно, как можно подойти к созданию правдоподобной и полной компьютерной модели мозга.

Параллельные компьютеры и «единственность» сознания

Многие считают, что развитие *параллельных* компьютеров содержит в себе ключ к построению машин, обладающих возможностями человеческого мозга. Далее мы кратко рассмотрим эту популярную сегодня идею.

Параллельный компьютер, в противоположность последовательному, может выполнять одновременно и независимо друг от друга огромное число отдельных операций, и результаты этих автономных операций время от времени объединяются, давая вклад в общий вычислительный процесс. Толчком к созданию такого типа компьютерной архитектуры послужили попытки моделирования нервной системы, поскольку, согласно современным представлениям, разные части мозга выполняют практически

автономные вычислительные функции (например, при обработке визуальной информации в зрительной коре).

Здесь необходимо сделать два замечания. Во-первых, между параллельным и последовательным компьютерами не существует *принципиальной* разницы. По сути, оба являются *машинами Тьюринга* (ср. главу 2, с. 74). Отличие может проявляться только лишь в эффективности, или скорости, вычислений в целом. Для некоторых типов вычислительных процессов параллельная организация, действительно, более эффективна, но это далеко не всегда так. Во-вторых, по крайней мере с моей точки зрения, крайне маловероятно, что классические параллельные вычисления дают ключ к тому, что происходит при *сознательном* мышлении. Характерным свойством сознательной мысли (по крайней мере в нормальном психологическом состоянии и не после операции по разделению полушарий мозга!) является ее «единственность» — в противоположность множественности выполняемых одновременно и независимо друг от друга операций.

Фразы типа: «Я же не могу думать обо всем сразу?!» можно услышать на каждом шагу. Можно ли *вообще* думать о нескольких вещах одновременно? Вероятно, кто-то *может* удерживать в голове несколько мыслей в одно и то же время, но это, скорее всего, будет похоже на постоянное перескакивание от одной мысли к другой и обратно, нежели на действительно одновременное, сознательное и независимое их обдумывание. Если бы кто-то мог думать о двух вещах совершенно независимо, то это было бы более похоже на обладание двумя *раздельными сознаниями*, пусть даже и на короткий промежуток времени, тогда как повседневный опыт (по крайней мере нормальных людей) свидетельствует о наличии *единственного* сознания, которое может иметь смутное представление о ряде вещей, но которое сконцентрировано в каждый момент времени только на *одной* из них.

Конечно, то, что мы подразумеваем здесь под «одной вещью», не совсем ясно. В следующей главе мы познакомимся с совершенно удивительными примерами «отдельных мыслей», появлявшихся в минуты вдохновения у Пуанкаре и Моцарта. Но нам вовсе не обязательно забираться так далеко, чтобы понять, что мысль человека в каждый конкретный момент времени может неявно быть очень сложной. Представьте себе, например, процесс обдумывания обеденного меню. Одна такая мысль может включать в себя такое количество разнообразной информации, что ее полное словесное описание было бы очень долгим.

«Единственность» осознанного восприятия представляется мне идущей вразрез с концепцией параллельного компьютера. С другой стороны, эта концепция может оказаться более подходящей в качестве модели *бессознательной* деятельности мозга. Различные независимые действия (ходьба, застегивание пуговиц, дыхание и даже разговор) могут выполняться человеком одновременно и более менее автономно, причем он может не осознавать *ни одно* из них!

С другой стороны, мне кажется, что эта «единственностью» сознания может иметь что-то общее с *квантовым параллелизмом*. Вспомним, что, согласно квантовой теории, на квантовом уровне различные альтернативы могут сосуществовать в линейной суперпозиции! Отсюда следует, что *одиночное квантовое состояние* могло бы, в принципе, состоять из большого числа различных событий, происходящих одновременно. Именно это и подразумевается под квантовым параллелизмом. Мы скоро рассмотрим теоретическую концепцию «квантового компьютера», в котором, в принципе, квантовый параллелизм мог бы быть использован для выполнения большого числа одновременных операций. Если «состояние ума», соответствующее рассудочной деятельности, имеет какое-то сходство с квантовым состоянием, то некая форма «единственности», или глобальности, мысли соответствует ему в большей степени, чем в случае обычного параллельного компьютера. У этой идеи есть несколько привлекательных аспектов, к которым я вернусь в следующей главе. Но прежде, чем рассматривать эту идею всерьез, мы должны ответить на вопрос, могут ли квантовые эффекты иметь какое-либо отношение к деятельности мозга.

Имеет ли квантовая механика отношение к работе мозга?

Все предыдущее обсуждение нервной деятельности проводилось целиком в рамках классических представлений, за исключением тех случаев, когда мы затрагивали физические явления, неявные причины которых отчасти обусловлены квантово-механическими эффектами (например, ионы, несущие единичные электрические заряды; натриевые и калиевые каналы; определенные химические потенциалы, определяющие триггерный характер генерации нервного импульса; химия нейромедиаторов). Но нет ли таких ключевых процессов в мозге, которые бы непосредственно определялись квантово-механическими эффектами? Для того чтобы рассуждения, описанные в конце предыдущей главы, имели какой-то смысл, такие процессы, по-видимому, должны существовать.

В действительности, можно указать, по крайней мере, одно место, где чисто квантовые явления имеют принципиальное значение для нервной деятельности, — это *сетчатая оболочка* глаза. (Вспомним, что сетчатка фактически входит в состав мозга!) Эксперименты с жабами показали, что в подходящих условиях адаптированная к темноте сетчатка вырабатывает макроскопический нервный импульс при попадании на нее *единичного фотона* (Бэйлор и др. [1979]). То же, как выясняется, справедливо и для человека (Хехт и др. [1941]), хотя в этом случае существует дополнительный механизм, который подавляет подобные слабые сигналы, тем самым очищая воспринимаемое изображение от лишнего визуального «шума». Необходимо суммарное воздействие примерно *семи* фотонов, чтобы адаптировавшийся к темноте испытуемый мог его ощутить. Тем не менее, в нашей сетчатке, по-видимому, все-таки есть клетки, чувствительные к попаданию только одного фотона.

Поскольку в теле человека *существуют* нейроны, способные срабатывать под воздействием единичного квантового события, то вполне обоснован вопрос о наличии таких клеток где-нибудь в основных отделах мозга. Насколько мне известно, это предположение не подтвердилось. У клеток всех изученных типов есть определенный порог срабатывания и требуется очень большое число квантов, чтобы перевести клетку в возбужденное состояние. Однако можно было бы допустить, что где-то глубоко внутри мозга должны быть клетки, чувствительные к одиночным квантам. Если это окажется верным, то квантовая механика должна играть существенную роль в деятельности мозга.

Но даже при таком положении вещей роль квантовой механики оказалась бы чисто номинальной, поскольку квант используется просто как возбудитель сигнала. Никаких интерференционных эффектов, характерных для квантовых явлений, пока обнаружить не удалось. Похоже, что в лучшем случае все, что мы можем получить от квантовой механики, это неопределенность момента срабатывания нейрона. Трудно представить, как это может пригодится нам на практике.

Однако некоторые вопросы, имеющие к этому отношение, не так тривиальны. Для их рассмотрения обратимся вновь к сетчатой оболочке глаза. Предположим, что фотон попадает на сетчатку, предварительно отразившись от полупрозрачного зеркала. Состояние фотона тогда будет представлять собой сложную линейную суперпозицию состояний, когда он попадает в клетку сетчатки и когда он проходит мимо клетки и вместо этого, скажем, улетает через окно в космос (см. рис. 6.17, с. 228) В тот момент, когда он *мог бы* попасть в клетку сетчатки, до тех пор, пока выполняется линейная процедура **U** (т. е детерминированная эволюция вектора состояния по уравнению Шредингера, см. с. 225), мы получим сложную линейную суперпозицию наличия и отсутствия нервного сигнала. Когда это доходит до сознания наблюдателя, воспринимается только *одна* из этих двух альтернатив, и должна использоваться другая квантовая **R**-процедура (редукция вектора состояния, см. с. 225). (Говоря так, я сознательно обхожу стороной теорию множественности миров, которая имеет множество своих собственных проблем!) В соответствии с рассуждениями, приведенными

в конце предыдущей главы, нам следует задать вопрос, достаточное ли количество материи вовлекается в прохождение сигнала, чтобы удовлетворялся *одногравитонный критерий* (см. главу 8)? Хотя при преобразовании энергии фотона в энергию движения массы при выработке сигнала в сетчатке достигается действительно гигантское усиление, возможно, до 10^{20} раз, эта масса все же значительно меньше величины планковской массы m_{Pl} (примерно в 10^8 раз). Однако нервный сигнал создает регистрируемое изменяющееся *электрическое поле* в окружающей среде (тороидальное поле с осью, совпадающей с нервным волокном, по которому оно перемещается). Это поле может вносить в *окружающую среду* значительное возмущение, за счет чего одногравитонный критерий будет легко удовлетворен. Таким образом, в соответствии с изложенной мной точкой зрения, **R**-процедура могла бы выполняться задолго до того, как мы увидим или, может случиться, не увидим вспышку света. К тому же, для редукции вектора состояния наше сознание не требуется!

Квантовые компьютеры

Если мы *все-таки* предположим, что чувствительные к одиночным квантам нейроны играют важную роль где-то в глубине нашего мозга, то возникает вопрос, какие следствия это могло бы иметь. Для начала я изложу концепцию *квантового компьютера*, предложенную Дойчем (см. также главу 4, с. 144), а затем мы выясним, можно ли ее рассматривать как имеющую отношение к теме нашей дискуссии.

Как было указано выше, главная идея состоит в использовании квантового параллелизма, в соответствии с которым два совершенно различных процесса должны рассматриваться как происходящие одновременно в виде квантовой линейной суперпозиции, например, фотон одновременно отражается от полупрозрачного зеркала и проходит через него или один и тот же фотон проходит через каждую из двух щелей. В случае квантового компьютера этими двумя различными наложенными друг на друга процессами будут два различных *вычисления*. При этом предполагается, что нас интересуют результаты не *обоих* вычислений, а некий результат, основанный на частичной информации, полученной из суперпозиции этих процессов. Наконец, когда оба вычисления завершены, над этой парой процессов должно быть проведено соответствующее «наблюдение», позволяющее получить искомый ответ [9]. Таким образом, это устройство могло бы сэкономить время за счет выполнения двух вычислений одновременно! До сих пор не видно никакого значительного преимущества от использования такого подхода, поскольку было бы гораздо проще непосредственно использовать два классических компьютера параллельно (или один классический параллельный компьютер), чем один квантовый. Однако реальные преимущества квантового компьютера могли бы проявиться при необходимости выполнить *очень большое*, возможно, неограниченно большое, количество параллельных вычислений, когда нас интересуют не их результаты сами по себе, а только подходящая комбинация результатов всех вычислений.

Принципиальное устройство квантового компьютера предполагает использование квантовой разновидности логических элементов, у которых выходной сигнал является результатом «унитарной операции» над входным сигналом — операции типа **U**, — и вся работа компьютера состояла бы в выполнении операции **U** до самого конца вычислений, пока конечный «акт наблюдения» не приведет к выполнению операции **R**.

Согласно выводам Дойча квантовые компьютеры не предназначены для выполнения неалгоритмических операций (т. е. действий, выходящих за пределы возможностей машины Тьюринга), но способны в некоторых, очень специфических случаях, достигать более высокого быстродействия (в смысле *теории сложности*, см. с. 144), чем обычная машина Тьюринга. Для такой блестящей идеи эти выводы представляются довольно неутешительными, но будем помнить о том, что пока мы стоим у самых истоков.

Какое отношение все это может иметь к работе мозга, содержащего значительное число нейронов, чувствительных к единич-

ным квантам? Провести аналогию здесь мешает в первую очередь то, что квантовые эффекты быстро теряются в «шуме» — мозг слишком «горяч», чтобы квантовая когерентность (поведение, которое удобно описывать как непрерывное действие U) сохранялась в нем сколько-нибудь продолжительное время. В моей терминологии это означало бы, что постоянно удовлетворяется одногравитонный критерий, так что операция R выполняется все время, изредка прерываясь операцией U.

Таким образом, пока у нас нет повода слишком надеяться на то, что квантовая механика откроет нам нечто новое о мозге. Возможно, мы все обречены быть просто компьютерами! Лично я в это не верю, но для окончательного выяснения вопроса нам необходимо идти дальше в наших исследованиях.

За пределами квантовой теории?

Я хочу вновь обратиться к вопросу, который проходит красной нитью через большую часть этой книги: действительно ли наши представления об окружающем мире, управляемом законами классической и квантовой физики в их современном понимании, адекватны для описания мозга и разума? «Обычное» квантовое описание нашего мозга определенно заходит в тупик, поскольку акт «наблюдения» считается важной составляющей правильной интерпретации общепринятой квантовой теории. Следует ли считать, что мозг «наблюдает сам себя» каждый раз при осознанном восприятии или возникновении мысли? Общепринятая теория не дает нам никаких указаний на то, каким образом квантовая механика могла бы принять это в расчет и, тем самым, как применить ее к мозгу в целом. Я попытался сформулировать вполне независимый от сознания критерий включения операции R («одногравитонный критерий»), и если нечто подобное удалось бы развить до полностью согласованной теории, то появилась бы возможность построения более ясного квантового описания мозга, чем существующее ныне.

Однако я считаю, что эти фундаментальные проблемы возникают не только при наших попытках описать деятельность мозга. Работа самих цифровых компьютеров существенно зависит от квантовых эффектов, пониманию которых, по-моему мнению, мешают трудности, внутренне присущие квантовой теории. Что это за «существенная» квантовая зависимость? Чтобы понять роль квантовой механики в цифровых вычислительных машинах, мы, прежде всего, должны выяснить, как можно заставить полностью *классический* объект вести себя подобно цифровому компьютеру. В главе 5 мы рассматривали классический «компьютер из биллиардных шаров» Фредкина—Тоффоли (с. 164); но, как мы видели, в этом теоретическом «устройстве» были использованы идеализации, позволяющие обойти проблему существенной нестабильности, внутренне присущей классическим системам. Эта проблема нестабильности, как указано выше (с. 173, рис. 5.14), проявляется в эффективном увеличении фазового объема эволюционирующей системы, которое почти неизбежно приводит к непрерывной потере точности операций, выполняемых классическим устройством. Именно квантовая механика позволяет в конце концов остановить это снижение точности. В современных электронных компьютерах необходимо существование *дискретных состояний* (скажем, для записи цифр **0** и **1**), всегда позволяющих однозначно установить, когда компьютер находится в одном, а когда в другом состоянии. Это выражает саму суть «цифровой» природы компьютерных операций. Эта дискретность, в конечном счете, достигается за счет квантовой механики. (Мы можем вспомнить здесь квантовую дискретность энергетических состояний, спектральных частот, значений спина и т. д., см. главу 6.) Даже старые механические вычислительные машины зависели от *прочности* различных своих частей, каковая, в свою очередь, непосредственно вытекает из дискретности квантовой теории [10].

Но квантовая дискретность не является только следствием операции U. Пожалуй, уравнение Шредингера в *еще меньшей степени* способно предотвратить нежелательное расплывание фазового объема и «потерю

точности», чем уравнения классической физики! Согласно **U**, волновая функция изолированной частицы, изначально локализованная в пространстве, будет все больше и больше расплываться с течением времени (с. 226). Если бы не действие **R** время от времени, более сложные системы тоже были бы подвержены такой беспричинной делокализации (вспомним кошку Шредингера). (*Дискретные* состояния атома, например, характеризуются определенными значениями энергии, импульса и полного момента импульса. Общее состояние, которое как раз «расплывается», представляет собой суперпозицию таких дискретных состояний. Именно процедура R на некотором этапе заставляет атом на самом деле «быть» в одном из этих дискретных состояний.)

Мне представляется, что ни классическая, ни квантовая механика — если только в последнюю не будут внесены дальнейшие фундаментальные изменения, которые превратили бы **R** в «реальный» процесс, — никогда не смогут объяснить механизм *мышления*. Возможно, что даже работа цифровых компьютеров требует более глубокого понимания взаимосвязи действий **U** и **R**. В случае с компьютерами, мы, по-крайней мере, знаем, что цифровые вычисления являются *алгоритмическими* (по самой конструкции!), и мы не пытаемся «обуздать» предполагаемую *неалгоритмичность* физических законов. Но я утверждаю, что в случае с мозгом и разумом ситуация совершенно иная. Вполне допустимо, что в процессе (сознательного) мышления участвует некая существенная неалгоритмическая составляющая. В следующей главе я попытаюсь подробно изложить причины, заставляющие меня верить в существование этой составляющей, а также выскажу предположения о том, какими удивительными реальными физическими эффектами обусловлено «сознание», влияющее на работу мозга.

Примечания

1. Из радиовещания BBC (см. Ходжис [1983], с. 419).
2. Первые эксперименты такого рода были проведены на кошках (см. Мирс, Сперри [1953]). За дальнейшими сведениями из области экспериментов с разделением полушарий мозга я отсылаю читателя к работам Сперри [1966], Газзаниги [1970] и МакКей [1987].
3. Доступное изложение принципов действия зрительной коры можно найти у Хьюбела [1988].
4. См. Хьюбел [1988], с. 221. Ранние эксперименты позволили обнаружить клетки, чувствительные только к образу руки.
5. Общепринятая сегодня теория, согласно которой нервная система состоит из отдельных клеток — нейронов — была впервые предложена и убедительно обоснована великим испанским нейрофизиологом Рамоном-и-Кахалом около 1900 года.
6. На самом деле, *любые* логические элементы могут быть построены с помощью одних только операций "~" и "&" (или даже только *одной*-единственной операции ~ (A&B)).
7. Фактически, использование логических элементов в большей степени отвечает конструкции электронного компьютера, чем изложенные в главе 2 особенности конструкции машины Тьюринга. В главе 2 особое внимание подходу Тьюринга было уделено по теоретическим соображениям. Начало действительному развитию компьютерных технологий положили в равной степени работы Алана Тьюринга и выдающегося американского математика венгерского происхождения Джона фон Неймана.
8. Эти сравнения во многом обманчивы. Подавляющее большинство транзисторов в современных компьютерах используется в устройствах «памяти» и не участвует в логических операциях; а память можно наращивать за счет внешних устройств практически бесконечно. При более интенсивном использовании параллельных вычислений количество транзисторов, непосредственно участвующих в выполнении логических операций, могло бы быть значительно больше, чем это принято в настоящее время.
9. Дойч в своих описаниях предпочитает использовать подход «множественности ми-

ров» относительно квантовой теории. Однако важно понимать, что это совершенно несущественно, поскольку концепция квантового компьютера принципиально не зависит от точки зрения на традиционную квантовую механику.

10. Этот комментарий перестает быть правомерным, если мы рассматриваем в качестве «классических» компонентов системы шестеренки, оси и т. п. Я предполагаю, что система состоит из обычных (скажем, точечных или сферических) частиц.

Глава 10

Где находится физика ума?

Для чего нужны умы?

В дискуссиях по проблеме «ум—тело» имеются два отдельных пункта, на которых обычно сосредоточивается внимание: «Каким образом материальный объект (мозг) может в действительности *пробуждать* сознание?»; и, наоборот: «Каким образом сознание усилием воли может реально воздействовать на (явно физически обусловленное) движение материальных объектов?» Это — пассивный и активный аспекты проблемы «ум—тело». Дело выглядит так, как если бы у нас в «уме» (или, вернее, в «сознании»), существовала некая нематериальная «вещь», которая, с одной стороны, активизируется материальным миром, а с другой — может оказывать на него воздействие. Я, однако, предпочитаю в своих предварительных замечаниях к этой последней главе обсудить до некоторой степени иной и, возможно, более научный вопрос, который относится к обеим проблемам, как пассивной, так и активной, — в надежде, что наши попытки найти ответ на него смогут приблизить нас к более глубокому пониманию этих извечных фундаментальных загадок философии. Мой вопрос звучит так: «Какое *преимущество естественного отбора* дает сознание тем, кто действительно им обладает?»

Такой формулировке вопроса присущи некоторые неявные допущения. Прежде всего — это уверенность в том, что сознание — это, на самом деле, «вещь», которую можно *научно описать*; что эта «вещь» действительно «что-то делает»; и, более того, что это «что-то» приносит пользу существам, которые им обладают, в то время как другие создания, подобные первым во всем, кроме наличия сознания, демонстрируют менее эффективное поведение. С другой стороны, можно полагать, что сознание — это лишь пассивный спутник достаточно совершенной системы управления, и само по себе, в действительности, *не* «делает» *ничего*. (Это последнее утверждение является, вероятно, точкой зрения сторонников «сильного» ИИ.) В качестве альтернативы можно рассмотреть иную концепцию, согласно которой существует некое божественное или таинственное предназначение сознания — быть может, носящее телеологический характер и нам пока не ведомое — так что любое обсуждение этого феномена только лишь в терминах естественного отбора неизбежно уведет нас в сторону от истины. С моей точки зрения, из всех доводов подобного толка, наиболее убедительно и наукообразно здесь выглядел бы так называемый *антропный принцип*, согласно которому природа нашей вселенной такова, потому что в ней в обязательном порядке требуется присутствие разумных существ-«наблюдателей» наподобие нас с вами. (Этот принцип был вкратце упомянут в главе 8, с. 308, и я еще вернусь к нему позже.)

Я собираюсь обсудить эти вопросы в должное время, но вначале мы должны заметить, что термин «ум», пожалуй, несколько уводит нас в сторону, когда мы говорим о проблеме «ум—тело». Ведь очень часто говорят о «бессознательном уме» и это

указывает на тот факт, что мы не рассматриваем термины «ум» и «сознание» как синонимы. Возможно, когда мы упоминаем о *бессознательном* уме, перед нами возникает неясный образ «суфлера», который незримо присутствует в каждой сцене, но кто обычно (кроме как, возможно, в снах, галлюцинациях, навязчивых состояниях или фрейдистских обмолвках) не посягает напрямую на контроль над нашим восприятием. Возможно, бессознательный ум в действительности *имеет* собственную способность осознавать, но в обычном состоянии это осознание пребывает совершенно отдельно от той части ума, которую мы традиционно называем «я».

Это, вообще говоря, не так уж и странно, как это может показаться на первый взгляд. Существуют эксперименты, которые, по-видимому, свидетельствуют о наличии определенного рода «сознания», присутствующего даже у пациента под общим наркозом на операционном столе — в том смысле, что разговоры, которые ведутся во время операции, могут быть впоследствии «неосознанно» восприняты пациентом или же быть «проявлены» позже под гипнозом как действительно «воспринимавшиеся» в прошлом. Более того, ощущения, которые, казалось бы, были вытеснены из сознания гипнотическим внушением, могут позднее быть выявлены во время другого сеанса гипноза как «уже пережитые», но каким-то образом оказавшиеся записанными «на другую дорожку» (см. Окли, Имз [1985]). Эти результаты мне не вполне ясны, хотя я и не думаю, что было бы правильно приписывать обычную способность осознания бессознательному уму, но у меня нет особого желания пускаться здесь в рассуждения по этим вопросам. Тем не менее, проведение различий между бессознательным и осознающим себя умом — это действительно сложная и тонкая тема, к которой нам еще придется вернуться.

Попытаемся достичь возможно большей ясности в описании того, что мы подразумеваем под «сознанием» и что считаем признаками его проявления. Я не думаю, что было бы умно на данной стадии понимания пытаться предлагать точное *определение* сознания, но мы можем в достаточной степени полагаться на наши субъективные впечатления и интуитивный здравый смысл относительно того, что этот термин означает, и когда описываемый им феномен проявляет себя. Мне более или менее понятно, когда я нахожусь в сознании, и склонен считать, что и другие люди испытывают при этом нечто подобное. Чтобы находиться в сознании, я должен, как мне кажется, осознавать *что-то*, может быть, такие ощущения, как боль или тепло, или красочный пейзаж, или звуки музыки; или, возможно, я осознаю такое чувство как изумление, отчаяние или счастье; или я могу осознавать воспоминание о некотором событии в прошлом, или начинаю понимать то, что говорит кто-то другой; или осознавать собственную новую идею; или я могу осознанно намереваться заговорить или предпринять какое-то другое действие, например, встать со стула. Я могу также «отстраниться» и осознавать подобные намерения, или мое ощущение боли, или опыт, запечатленный в памяти, или акт понимания; или я могу даже просто осознавать свое собственное сознание. Я могу находиться в состоянии сна и все равно быть до некоторой степени осознающим происходящее, если мне снится сон; или, возможно, когда я начинаю просыпаться, я сознательно воздействую на развитие этого сна. Я готов считать, что сознание — это нечто, имеющее некоторую градацию, а не просто что-то, что есть или чего нет. Я считаю слово «сознание» в сущности синонимичным слову «осознание» (хотя, возможно, «осознание» немного пассивнее, чем то, что я понимаю под «сознанием»), в то время как «ум» и «душа» имеют дополнительные оттенки смысла, которые в значительной мере *менее* отчетливо определимы в настоящее время. У нас будет много хлопот с пониманием того, что такое «сознание» само по себе, поэтому я надеюсь, что читатель меня простит, если я оставлю в покое дальнейшие проблемы, связанные с терминами «ум» и «душа»!

Существует также вопрос о том, что подразумевать под словом *интеллект*. В конце концов, именно об этом объекте — а не о более расплывчатом понятии «сознания» — предпочитают говорить люди, связанные с ИИ. Алан Тьюринг в своей знаменитой работе (Тьюринг [1950]) (см. главу 1, с. 41)

рассматривал непосредственно не столько «сознание», сколько «мышление», а слово «интеллект» даже было вынесено им в заглавие. На мой взгляд, вопрос об интеллекте является вторичным по отношению к вопросу о феномене сознания. Едва ли я поверю в то, что настоящий интеллект мог бы действительно существовать, когда бы его не сопровождало сознание. С другой стороны, если в итоге и вправду окажется, что приверженцы ИИ *способны* моделировать интеллект без присутствия сознания, *тогда было бы совершенно неудовлетворительным* определять интеллект, не включая в это понятие такой моделированный интеллект. Но в этом случае «интеллект» как предмет обсуждения оказался бы вне поля моего внимания, поскольку мой интерес связан, в первую очередь, с «сознанием».

Когда я высказываю свое убеждение, что истинный интеллект требует присутствия сознания, я при этом неявно предполагаю (поскольку я не разделяю точку зрения сторонников теории «сильного» ИИ, согласно которой простое применение алгоритма способно пробуждать сознание), что интеллект не может надлежащим образом моделироваться алгоритмическими средствами, то есть путем использования компьютера так, как это делается сегодня. (См. обсуждение «теста Тьюринга» в главе 1.) Очень скоро (см., в частности, обсуждение математического мышления, приведенное тремя разделами ниже, на с. 356) я постараюсь привести самые убедительные доводы в пользу необходимости присутствия существенно *неалгоритмической* составляющей в работе сознания.

Теперь обратимся к вопросу о том, *существует ли* четкое различие между одним объектом, который обладает сознанием, — и другим, «эквивалентным» первому во всем, кроме способности сознавать. Всегда ли сознание, присущее некоторому объекту, проявляет свое присутствие? Я предпочитаю думать, что ответить на этот вопрос следует однозначно «да». Однако, эта моя вера едва ли найдет поддержку в научных кругах, если там до сих пор нет согласия даже в вопросе о том, где можно найти сознание в царстве животных. Некоторые вообще не допускают мысли, что им могут обладать какие бы то ни было животные, отличные от людей (а некоторые придерживаются того же мнения и в отношении человеческих существ, живших за 1000 или более лет до н. э.; см. Джейнс [1980]); и в то же время кто-то готов допустить наличие сознания у насекомых, у червей и даже — почему бы нет? — у камней! Что касается меня, то я склонен сомневаться в том, что червь или насекомое — не говоря уже о камнях — в значительной степени (если вообще) обладают этим качеством; но млекопитающие, в общем и целом, подчас производят на меня впечатление существ, способных на подлинное осознание. Имея столь диаметрально противоположные точки зрения, приходится констатировать, что на сегодняшний день общепринятый критерий проявления сознания отсутствует. Правда, вполне возможно, что *есть все же* критерий сознательного поведения, хотя он и не заслужил всеобщего признания. Но не вызывает сомнений, что в любом случае только *активная* роль сознания могла бы иметь принципиальное значение, поскольку невозможно представить себе, чтобы простое наличие способности осознавать, без активного дополнения к ней, может быть непосредственно зафиксировано. Подтверждением этому факту послужили ужасные случаи применения в 40-е годы лекарства на основе яда кураре в качестве «анестезирующего» средства при операциях, проводимых на маленьких детях, — тогда как действительный эффект этого средства заключается в парализации воздействия двигательных нервов на мускулы, из-за чего агония, которую в буквальном смысле *испытывали* несчастные дети, оставалась на протяжении операции незаметной для хирурга (см. Деннетт [1978], с. 209).

Вернемся к той гипотетической активной роли, которую *может* иметь сознание. Верно ли, что сознание может играть — а часто и играет — активную операционально различимую роль? Я полагаю, что это должно быть так и постараюсь сейчас обосновать свою убежденность несколькими независимыми доводами. Во-первых, благодаря нашему «здравому смыслу» мы часто ощущаем, что мы непосредственно воспринимаем, что другой человек *находится* в сознании. *Такое* впечатление вряд ли может

быть ошибочным[1]. В то время как человек, который *находится* в сознании, может (подобно детям под действием кураре) и не подавать соответствующих признаков — находящийся в *бессознательном* состоянии едва ли будет выглядеть как человек, обладающий сознанием! Следовательно, должен существовать некий тип поведения, который можно было бы назвать характерным для человека, пребывающего в сознании (хотя даже и *не всегда* подтверждаемый самим сознанием), который мы бы воспринимали именно так благодаря нашим «интуитивным представлениям».

Во-вторых, примем во внимание безжалостный процесс естественного отбора. Будем рассматривать этот процесс в свете того факта, что, как мы видели в предыдущей главе, не вся активность мозга непосредственно доступна сознанию. И действительно, более «древний» мозжечок — обладающий значительным (по сравнению с остальными частями головного мозга) превосходством в плотности нейронов — производит, по-видимому, весьма сложные действия безо всякого вмешательства со стороны сознания. Однако, природа избрала для эволюционного развития таких сознающих себя и окружающий мир существ, как мы, вместо того, чтобы удовлетвориться созданиями, которые вполне могли бы существовать при помощи абсолютно бессознательных механизмов управления. Если сознание не служит целям селекции, то зачем природа занялась созданием *«сознательных» разновидностей* мозга, тогда как не наделенные сознанием «мозги-автоматы», наподобие мозжечка, могли бы функционировать не менее успешно?

Более того, существует простая «основополагающая» причина для предположения о том, что сознание должно иметь *какое-то* активное влияние, даже если его результат *не* является преимуществом при естественном отборе. Ибо почему еще мы (или существа, нам подобные) можем иной раз мучиться при попытке ответить на вопрос — особенно, если изучается эта тема — «о самих себе». (Мне так и хочется сказать: «Почему *вы* читаете эту главу?» или «Почему *у меня* было сильное желание написать книгу именно на эту тему?») Трудно себе представить, чтобы полностью лишенный сознания автомат стал бы тратить время на подобные вещи. А поскольку обладающие сознанием существа, с другой стороны, время от времени поступают *как раз таким* вот смешным образом, то их поведение *отличается* от поведения остальных — откуда следует, что сознание все-таки производит *определенное* активное воздействие! Разумеется, не составит труда специально запрограммировать компьютер так, чтобы он вел себя столь же нелепым образом (например, он мог бы в согласии с заложенным в него алгоритмом постоянно повторять на ходу: «О Господи, ну в чем же смысл жизни? Почему я здесь нахожусь? Что такое, черт побери, это „Я", которым я себя ощущаю?)». Но почему же естественный отбор позаботился о создании благоприятных условий именно для такой расы индивидов, когда жестокий закон джунглей наверняка давно бы выдрал с корнем подобную бесполезную ерунду!

Мне кажется очевидным, что все эти размышления и бормотание, которым мы (временно становясь философами) предаемся, не могли быть самоценными для процесса естественного отбора, а являются просто необходимым с его (естественного отбора) точки зрения «багажом», который должны нести существа, *обладающие* подлинным сознанием — при том, что само оно возникло в ходе естественного отбора по совершенно другой и, вероятно, очень серьезной причине. Этот багаж не слишком обременителен, и легко (хотя, быть может, и неохотно) переносится, скорее всего, именно непреклонными силами естественного отбора. В тех случаях, когда на земле царят мир и процветание, которыми человеческий род время от времени имеет счастье наслаждаться (ибо нам не всегда приходится бороться со стихиями (или нашими соседями) за выживание) — тогда, возможно, сокровища, содержащиеся в нашем багаже, становятся предметом удивления и любопытства. Именно в такие моменты, глядя на окружающих тебя философствующих людей, всерьез *убеждаешься* в том, что они, как и ты, тоже обладают умом.

---

[1] По крайней мере, при наличии современных компьютерных технологий (см. обсуждение теста Тьюринга в главе 1).

Что в действительности делает сознание?

Давайте согласимся с тем, что наличие у данного существа сознания является его реальным преимуществом в ходе естественного отбора. В чем конкретно может заключаться это преимущество? Я знаком с одной точкой зрения, согласно которой способность осознавать происходящее может оказаться полезной хищнику в его попытке предугадать, что его жертва будет делать в следующий момент, с помощью мысленной «постановки себя на место» этой жертвы. *Воображая* себя своей жертвой, он мог бы увеличить шансы удачного исхода охоты.

Вполне может быть, что эта идея отчасти верна, но кое-что в ней меня смущает. Во-первых, здесь предполагается наличие некоторого сознания у самой жертвы, так как вряд ли имело бы смысл представлять себя автоматом, поскольку автомат — по определению, *не обладающий сознанием* — ни в коем случае не есть то, чем *можно* «быть»! В любом случае, я мог бы с таким же успехом представить себе, что совершенно не обладающий сознанием автомат-хищник имеет внутри в качестве подпрограммы последовательность действий, которая в точности соответствовала бы действиям автомата-жертвы. Мне вообще не кажется, что необходимость вводить сознание в отношения типа «хищник—жертва» может быть *логически* обоснована.

Разумеется, трудно понять, как случайные процессы естественного отбора могли быть достаточно умными, чтобы дать *автомату*-хищнику полную копию программы жертвы. Это бы выглядело скорее как *шпионаж*, а не как естественный отбор! А *частичная* программа (в смысле отрезка «ленты» машины Тьюринга или чего-то подобного такой ленте) вряд ли была бы достаточно полезна хищнику с точки зрения естественного отбора. Ему был бы необходим какой-нибудь завершенный фрагмент этой ленты, а лучше (хотя и менее вероятно) — вся лента целиком. Следовательно (и как некая альтернатива этому), частичная правда могла бы заключаться в том, что модель поведения «хищник—жертва» подразумевает все же наличие не компьютерной программы, но определенного элемента сознания. Однако, это, по-моему, уже не относится к *настоящему* вопросу о различии между сознательным и «запрограммированным» действиями.

Суть идеи, о которой шла речь выше, по-видимому, напрямую связана с широко распространенной точкой зрения, согласно которой считается, что система может «осознавать» нечто, если в ней уже имеется модель этого объекта, и, соответственно, что она становится *«самосознающей»*, когда она в *самой себе* содержит модель *самой себя*. Но если компьютерная программа содержит (скажем, в качестве подпрограммы) текст другой компьютерной программы, то первая при этом не может осознавать вторую, также как и не может приводить к самосознанию обращение программы к себе самой. Несмотря на подобные часто встречающиеся заявления, фундаментальные понятия самосознания и способности к осознанию в этих рассуждениях едва ли затрагиваются. Видеокамера не осознает сцены, которые она снимает; как не обладает самосознанием и видеокамера, направленная на зеркало (рис. 10.1).

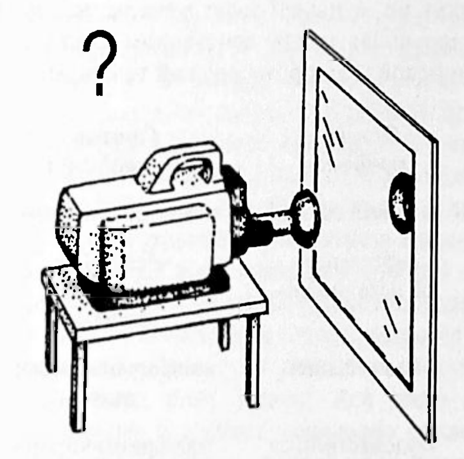


Рис. 10.1. Видеокамера, направленная на зеркало, строит модель себя внутри самой себя. Становится ли она от этого самосознающей?

Я хочу пойти в другом направлении. Как мы уже знаем, не вся деятельность нашего мозга сопровождается работой сознания (в особенности, это относится к действиям, которые управляются мозжечком). А какая часть из того, что мы можем делать

осознанно, не может быть сделана бессознательно? Проблема становится еще менее ясной из-за того, что все, для чего нам изначально требуется сознание, похоже, может быть со временем заучено и впоследствии выполняться уже автоматически (возможно, именно мозжечком). Кажется, что сознание требуется, чтобы справляться с ситуациями, где нам приходится высказывать новые суждения, и где правила не были заданы заранее. Трудно достичь большой точности при проведении различий между теми видами умственной деятельности, которые, по-видимому, требуют подключения сознания, и теми, которые нет. Сторонники «сильного» ИИ (да и многие другие) будут настаивать на том, что «формирование новых суждений» — это не более, чем повторное применение ряда точно сформулированных алгоритмических правил — только теперь на некоем полумистическом «высоком уровне», так, что их действия нами при этом не осознаются. Однако, как мне кажется, даже в нашем обиходном лексиконе есть такие термины — используемые нами в повседневной жизни для разделения умственной деятельности на осознанную и бессознательную — которые уже сами по себе могут по меньшей мере *навести на мысль* о различиях между действиями неалгоритмической и алгоритмической природы:

Сознание требуется	*Сознание не требуется*
«здравый смысл»	«автоматический»
«суждение об истинности»	«бездумное следование правилам»
«понимание»	«запрограммированный»
«художественная оценка»	«алгоритмический»

Возможно, эти различия не всегда достаточно четко очерчены хотя бы потому, что в наши сознательные суждения входит немало *неосознанных* факторов: опыт, интуиция, предрассудки, даже наше привычное использование логики. Но я утверждаю, что сами по себе суждения — это проявления работы *сознания*. Поэтому я полагаю, что, тогда как бессознательные действия мозга происходят в соответствии с алгоритмическими процессами, действие сознания имеет совершенно иную природу, и потому не может быть описано никаким алгоритмом.

Есть некая ирония в том, что многое из того, что я излагаю здесь, является полной противоположностью по отношению к некоторым другим точкам зрения, которые мне приходится довольно часто слышать. Например, утверждают, что именно *сознательный* ум «рационален» и доступен пониманию; тогда как бессознательные действия нередко загадочны и труднообъяснимы. Те, кто работает с искусственным интеллектом, часто считают, что как только мы сможем понять ход осознанной мысли, то можно сразу же будет придумать соответствующий алгоритм для его компьютерной реализации; а вот таинственные бессознательные процессы нашему пониманию (пока!) не доступны. В моем представлении эти процессы вполне могут быть алгоритмическими, но при этом настолько сложными, что их детальный анализ практически невозможен. Четко осознаваемый ход мысли, который может быть разумно объяснен как нечто полностью логичное, в свою очередь может быть (зачастую) переведен на язык алгоритмов — но на *совершенно ином уровне*. Мы сейчас говорим не о внутренних процессах (возбуждении нейронов и т. п.), а о манипулировании законченными мыслями. Иногда оно носит алгоритмический характер (как в случае традиционной логики: древнегреческих силлогизмов, формализованных Аристотелем; или символьной логики, разработанной математиком Джорджем Булем; см. Гарднер [1958]); а иногда — неалгоритмический (как в случае с теоремой Геделя или некоторыми примерами, приведенными в главе 4). А как реализовать на компьютере *формирование суждений*, которое я рассматриваю как критерий наличия сознания, — об *этом* разработчики ИИ не имеют даже ни малейшего представления!

Иногда мне возражают, что, поскольку *критерии* для этих суждений не являются в конце концов осознанными, то как я могу приписывать такие суждения сознанию? Однако, тем самым мои оппоненты упускают самую суть тех идей, которые я пытаюсь

выразить. Ведь я не требую, чтобы мы осознавали, *как* мы формируем наши сознательные впечатления и суждения. Это привело бы к смешению тех уровней, о которых я только что упоминал. Истинные *основания* наших осознанных впечатлений наверняка будут недоступны сознанию. Они должны были бы рассматриваться на более глубоком (материальном) уровне по сравнению с первопричинами наших явных мыслей, которые мы непосредственно осознаем. (Ниже я осмелюсь предложить на рассмотрение одну интересную гипотезу на этот счет!) Собственно сознательные впечатления и *являются* (неалгоритмическими) суждениями.

Эта тема, затрагивающая вопрос о возможной *неалгоритмической* составляющей механизма нашего осознанного мышления, проходила красной нитью и через все предыдущие главы. В частности, заключительная часть дискуссии в главе 4, особенно касающаяся теоремы Геделя, подводила к мысли о том, что (по крайней мере, в математике) сознательное «вглядывание» иной раз позволяет нам устанавливать справедливость утверждения способом, недоступным для алгоритма. (Подробнее я остановлюсь на этом доводе чуть позже.) Конечно же, сами по себе алгоритмы не способны находить истину! Построить алгоритм, генерирующий только ложные суждения, столь же просто, как и алгоритм, результатом работы которого были бы одни только истины. Для определения пригодности или непригодности того или иного алгоритма нам совершенно необходимо своего рода вдохновение, интуитивное прозрение, приходящее извне (далее я еще вернусь к этому вопросу). И я утверждаю, что именно эта способность провидения (или «интуитивного постижения») глубокого различия между истиной и ложью (равно как и между красотой и уродством!) является признаком наличия сознания.

Я, однако, должен сразу же оговориться, что ни в коем случае не имею здесь в виду какое-то мистическое «ясновидение». Сознание абсолютно бесполезно при попытке угадать счастливое число в (честно проводимой) лотерее! Я имею в виду суждения, которые постоянно формируются человеком в сознательном состоянии, когда собираются воедино и сопоставляются все относящиеся к предмету размышлений факты, данные чувственного опыта, воспоминания — а в иную минуту вдохновения даже рождаются мудрые мысли. В принципе, мы располагаем достаточным количеством информации для того, чтобы вынести соответствующее суждение — но процесс его осмысленного формирования путем выделения необходимой информации из трясины фактов может просто не иметь точного выражения на языке алгоритмов (или же подобное выражение существует, но может оказаться при этом бесполезным практически). Возможно, мы находимся в ситуации, что когда суждение уже сделано, некоторый алгоритмический процесс (или просто более простое суждение) проверяет его справедливость, но не его изначальное формирование. В такой ситуации, как мне кажется, сознание «нашло бы себя» в роли создателя подходящих суждений.

Почему я утверждаю, что неалгоритмическое построение суждений является критерием наличия сознания? Отчасти я опираюсь здесь на свой опыт ученого-математика. Я просто не доверяю своим механическим действиям, если они не были сперва придирчиво исследованы сознанием. Часто сам по себе алгоритм, использующийся при определенных вычислениях, не вызывает сомнения — но *тот ли* алгоритм используется для решения данной конкретной задачи? Рассмотрим простой пример: если вас заставят вызубрить алгоритмы перемножения двух чисел и деления одного числа на другое (или даже разрешат использовать запрограммированный карманный калькулятор) — гарантирует ли это, что вы сможете определить в каждом конкретном случае, какое из этих действий приведет к решению поставленной перед вами задачи? Для этого нужно *думать* и строить *осознанное* суждение. (Вскоре мы увидим, почему такие суждения должны быть, по крайней мере иногда, *не*алгоритмическими!) Разумеется, коль скоро вы решите большое количество однотипных задач, выбор между умножением и делением станет настолько привычным, что будет выполняться совершенно автоматически — не исключено, что при участии одного лишь мозжечка. На этой стадии осознанное восприятие происходящего не является

больше необходимым, поэтому можно спокойно позволить своему сознанию занять ум иными проблемами или просто «отпустить его в свободное плавание» — разве что время от времени проверяя ход выполнения алгоритма.

То же самое постоянно происходит на всех уровнях математического мышления. Люди часто стремятся найти адекватные алгоритмы, когда занимаются математикой, но само это стремление отнюдь не кажется алгоритмической процедурой. Как только подходящий алгоритм найден, задача, в некотором смысле, уже решена. Более того, определение с точки зрения математики степени точности или пригодности алгоритма требует значительных усилий со стороны сознания. Нечто подобное имело место при обсуждении формальных систем для математики, которые были описаны в главе 4. Если начать с формулировки нескольких аксиом, то затем из них можно вывести различные математические утверждения. Не исключено, что последняя операция может оказаться алгоритмической, но все же изначально математик должен осознанно решить вопрос об адекватности этих аксиом. Почему это решение с необходимостью будет *не*-алгоритмическим, должно стать ясным из рассуждений, идущих непосредственно после следующего параграфа. Но прежде, чем мы перейдем к этому вопросу, давайте посмотрим, какая теория возникновения мозга и принципов его деятельности является на сегодняшний день наиболее популярной.

Естественный отбор алгоритмов?

Если предположить, что умственная деятельность человека — как осознанная, так и нет — это всего лишь выполнение очень сложного алгоритма, то сразу же возникает вопрос: а как, собственно, мог возникнуть такой в высшей степени эффективный алгоритм. Стандартным ответом здесь, разумеется, будет «естественный отбор». Как только стали появляться существа, наделенные мозгом, между ними возникла конкуренция, в которой побеждали (т. е. выживали и производили более многочисленное потомство) те, чей алгоритм оказывался эффективнее. Их потомки тоже имели, как правило, более эффективные алгоритмы поведения, чем их родственники, коль скоро им посчастливилось унаследовать составляющие этих «продвинутых» алгоритмов от своих родителей; так постепенно алгоритмы улучшались — не обязательно стабильно, поскольку могли случаться и значительные колебания в их эволюции — пока не было достигнуто то поразительное совершенство, которое (по всей видимости) можно обнаружить в мозге современного человека (см. Доукинс [1986]).

Даже если судить сообразно моей точке зрения, *некоторая* доля истины в этой картине должна быть, поскольку, как мне представляется, большая часть работы мозга действительно носит алгоритмический характер, и к тому же — как читатель наверняка догадался из предыдущих рассуждений — я являюсь убежденным сторонником (теории) естественного отбора. Но я не понимаю, как естественный отбор сам по себе мог дать рождение алгоритмам, которые позволяли бы делать осознанные выводы касательно *правомерности* применения всех прочих алгоритмов, которыми мы должны, по идее, пользоваться.

Представьте себе обычную компьютерную программу. Как *она* появилась на свет? Ясно, что никак не за счет (непосредственно) естественного отбора! Чтобы это произошло, какой-нибудь программист должен был бы разработать ее и убедиться, что она корректно выполняет те действия, для которых она предназначена. (В действительности большинство сложных компьютерных программ содержат ошибки — как правило, незначительные и малозаметные, которые выявляются только в достаточно редких случаях при необычных стечениях обстоятельств. Наличие таких ошибок не влияет существенно на мои рассуждения.) Иногда компьютерная программа может быть «написана» другой компьютерной программой-«мастером», но тогда та, в свою очередь, с необходимостью должна быть создана человеческим гением; то же самое относится и к тем программам, в состав которых могут входить фрагменты кодов, написанных другими компьютерными программами. Но в любом случае, задача обоснования

использования конкретного алгоритма и разработка общей концепции программы «ложится на плечи» (по крайней мере) одного человеческого сознания.

Можно представить себе, конечно, что все могло бы происходить совсем не так, и что по прошествии достаточного количества времени компьютерная программа могла бы самостоятельно эволюционировать в процессе некоего естественного отбора. Если верить, что действия, производимые в сознании программистов, сами являются всего лишь алгоритмами, то надо с необходимостью согласиться с тем, что таким же образом *развились* и алгоритмы. Но меня здесь беспокоит то, что принятие решения о правомерности использования алгоритма отнюдь не является алгоритмическим процессом. Кое-что об этом уже было сказано в главе 2. (Будет или нет машина Тьюринга на самом деле *останавливаться* — вопрос, который не может быть решен алгоритмическим путем.) Чтобы решить, будет ли алгоритм действительно *работать*, нужно *глубокое понимание*, а не просто еще один алгоритм.

Тем не менее, можно было бы представить себе, что существует определенный процесс естественного отбора, который способен создавать *довольно* эффективные алгоритмы. Лично мне, однако, очень трудно в это поверить. Любой процесс отбора такого рода мог бы оказывать воздействие только на *результаты* выполнения алгоритмов [2], а не на лежащие в основе этих алгоритмов идеи. И это не только совершенно неэффективно — я думаю, что это не принесло бы вообще никакого результата. Во-первых, нелегко определить, глядя на итог работы алгоритма, что он из себя на самом деле представляет. (Нетрудно сконструировать две простые, но совершенно различные процедуры для машины Тьюринга, выходные ленты которых совпадали бы, скажем, до 2^{65536}-й позиции — и тогда их различие не было бы замечено даже за всю историю вселенной!) Более того, малейшая «мутация» алгоритма (например, небольшое изменение в описании машины Тьюринга или в ее выходной ленте) сделала бы ее полностью бесполезной; поэтому трудно понять, как настоящие *усовершенствования* алгоритмов могут получаться таким вот случайным образом. (Даже *обдуманные* усовершенствования труднореализуемы, когда неизвестен их точный «смысл». Так традиционно получается в тех нередко возникающих ситуациях, когда необходимо внести изменения или исправления в сложную и небрежно задокументированную программу, чей автор находится вне пределов досягаемости или давно умер. Тогда вместо того, чтобы пытаться разобраться в хитросплетениях разнообразных промежуточных значений и неявных подзадач, на которых базируется эта программа, иной раз бывает проще стереть все и начать заново!)

Предположим, что мог бы быть разработан гораздо более «здоровый» метод определения алгоритмов, в отношении которого вышеприведенная критика становилась бы беспочвенной. В некотором роде, это как раз то, о чем я и говорю. «Здоровые» определения — это *идеи*, на которых базируется алгоритм. Но идеям, насколько нам известно, для своего выражения требуется разум, наделенный сознанием. А значит, мы вновь возвращаемся к проблеме определения сознания и тех его свойств, которые отличают обладающих сознанием существ от остального мира — и к вопросу о том, как, черт побери, естественный отбор мог оказаться достаточно «умным», чтобы развивать *такие замечательные* свойства.

Результаты естественного отбора и в самом деле удивительны. Те скромные познания о функционировании человеческого мозга — и, разумеется, мозга любых других существ — которыми я обладаю, просто ошеломляют меня, заставляя испытывать благоговейный трепет. Работа отдельного нейрона поразительна, но все вместе нейроны представляют еще более впечатляющую структуру, с самого момента рождения насчитывающую огромное множество соединений и связей, которые в дальнейшем позволят ей решать практически любые задачи. Замечательно не только само

[2] Здесь можно упомянуть еще один непростой вопрос относительно того, могут ли два алгоритма рассматриваться как эквивалентные друг другу, если *результаты* их действий — но не сами вычисления! — являются тождественными. См. главу 2, с. 79.

сознание, но и все то, что связано с обеспечением его деятельности!

Если бы однажды нам довелось обнаружить то свойство, которое позволяет физическому объекту обретать сознание, то, изучив его во всех деталях, мы, вероятно, могли бы сконструировать подобные объекты для собственных нужд — хотя они не обязательно расценивались бы как «машины», в современном понимании этого слова. Нетрудно предположить, что такие объекты могли бы многократно нас превосходить, ибо они были разработаны *специально* для этой задачи — т. е. для *обретения сознания*. Им бы не пришлось вырастать из одной клетки. Им бы не пришлось нести на себе «багаж» предков (старые и «бесполезные» части мозга или тела, которые продолжают существовать в нас только благодаря «несчастьям», приключившимся с нашими далекими предками). Можно также представить себе, что благодаря этим преимуществам, такие объекты могли бы *по-настоящему* заменять собой людей там, где (по мнению тех, кто согласен со мной) алгоритмические компьютеры обречены на выполнение обслуживающих функций.

Но вполне возможно, что тема сознания имеет гораздо больше аспектов. Может быть и так, что каким-то образом наше сознание действительно зависит от нашего наследия и от миллионов лет эволюции, лежащих у нас за спиной. Меня не покидает ощущение, что в самой эволюции, в ее явном «нащупывании» пути к какой-то будущей цели есть что-то загадочное и непостижимое. Кажется, что все организовано несколько лучше, чем оно «должно было быть» на основе слепой эволюции и естественного отбора. Вполне возможно, однако, что внешние проявления здесь обманчивы. Возможно, это как-то связано с тем способом, каким действуют физические законы, что позволяет естественному отбору протекать гораздо эффективнее, чем в случае, если бы этот процесс управлялся произвольными законами. Возникающее в результате явно «интеллектуальное нащупывание» — это отдельная интересная тема, к которой я вернусь несколько позже.

Неалгоритмическая природа математической интуиции

Как я уже указывал ранее, моя уверенность в том, что сознание способно влиять на характер суждений об истинности *не*-алгоритмическим путем, опирается главным образом на результаты теоремы Геделя. Если мы видим, что сознание действует неалгоритмически при формулировании *математических* суждений, где вычисления и строгие доказательства являются непременным требованием, то уж наверняка нас нетрудно будет убедить и в том, что эта неалгоритмическая составляющая могла бы являться решающей и для роли сознания при более общих (не связанных с математикой) обстоятельствах.

Вспомним доводы, приведенные в главе 4 в рамках доказательства теоремы Геделя и устанавливающие ее применимость к решению вопроса о вычислимости. Там было показано, что *какой бы* (достаточно сложный) алгоритм ни использовал математик для установления математической истины или, что то же самое [1], какую бы *формальную систему* он[3)] ни принял для задания своего критерия истинности — всегда найдутся математические суждения, подобные сформулированному Геделем утверждению $P_k(k)$ для системы (см. с. 118), на которые его алгоритм не сможет дать ответа. Если ум математика работает полностью алгоритмически, то алгоритм (или формальная система), которые он обычно использует для построения своих суждений, оказываются не в состоянии справиться с утверждением $P_k(k)$, полученным с помощью его собственного алгоритма. Тем не менее, *мы* можем (в принципе) понять, что $P_k(k)$ на самом деле *истинно*! Этот факт, по всей видимости, должен был бы указать *ему* на противоречие, поскольку *он*, как и мы, не может не заметить его. А это, в свою очередь, может свидетельствовать о *не*-алгоритмическом характере его рассуждений!

В этом заключается суть довода, предложенного Лукасом [1961] в поддержку точ-

3) Разумеется, «он» означает «она или он». См. сноску на с. 42.

ки зрения, согласно которой деятельность мозга не может быть полностью алгоритмической, против которого, однако, время от времени выдвигались различные контрдоводы (см., например, Бенасерраф [1967], Гуд [1969], Льюис [1969, 1989], Хофштадтер [1981], Бови [1982]). В связи с этой дискуссией я должен подчеркнуть, что термины «алгоритм» и «алгоритмический» относятся к чему угодно, что может быть (достоверно) смоделировано на компьютере общего назначения. Сюда включается, конечно, как «параллельная обработка», так и «нейросети» (или «машины с переменной структурой связей»), «эвристика», «обучение» (где всегда заранее задается определенный фиксированный шаблон, по которому машина должна обучаться), а также взаимодействие с внешним миром (которое может моделироваться посредством входной ленты машины Тьюринга). Наиболее серьезным из этих контраргументов является следующий: чтобы *действительно* убедиться в истинности утверждения $P_k(k)$, нам нужно *знать*, какой именно алгоритм использует математик, и при этом быть уверенным в правомерности его использования в качестве средства достижения математической истины.

Если в голове у математика выполняется очень сложный алгоритм, то у нас *не будет возможности* узнать, что он из себя представляет, и поэтому мы не сможем сконструировать для него утверждение геделевского типа, не говоря уже об уверенности в обоснованности его применения.

Такого типа возражения часто выдвигаются против утверждений подобных тому, которое я привел в начале этого раздела, а именно, что теорема Геделя свидетельствует о неалгоритмическом характере наших математических суждений. Но сам я не нахожу это возражение слишком убедительным. Предположим на мгновение, что способы, которыми математики формируют осознанные суждения о математической истине действительно *являются* алгоритмическими. Попробуем, используя теорему Геделя, доказать абсурдность этого утверждения от противного (*reductio ad absurdum*!).

Прежде всего мы должны рассмотреть возможность того, что разные математики используют *неэквивалентные* алгоритмы для суждения об истинности того или иного утверждения. Однако — и это одно из наиболее поразительных свойств математики (может быть, почти единственной в этом отношении среди всех прочих наук) — истинность математических утверждений может быть установлена посредством абстрактных рассуждений! Математические рассуждения, которые убеждают одного математика, с необходимостью убедят и другого (при условии, что в них нет ошибок и суть нигде не упущена). Это относится и к утверждениям типа геделевского. Если первый математик готов согласиться с тем, что все аксиомы и операции некоторой формальной системы всегда приводят только к *истинным* утверждениям, то он также должен быть готов принять в качестве истинного и соответствующее этой системе геделевское утверждение. Точно то же самое произойдет и со вторым математиком. Таким образом, рассуждения, устанавливающие математическую истину, являются *передаваемыми* [2].

Отсюда следует, что мы, говоря об алгоритмах, имеем в виду не какие-то неясные разномастные построения, которые, возможно, рождаются и бродят в голове каждого отдельного математика, а *одну* универсально применяемую формальную систему, которая *эквивалентна всем* возможным алгоритмам, использующимся математиками для суждений о математической истине. Однако мы никак не можем знать, является ли эта гипотетическая «универсальная» система той, которая используется математиками для установления истинности. Ибо в этом случае мы *могли бы* построить для нее геделевское утверждение, и знали бы наверняка, что оно математически истинно. Следовательно, мы приходим к заключению, что алгоритм, который математики используют для определения математической истины, настолько сложен или невразумителен, что даже правомерность его применения навсегда останется для нас под вопросом.

Но это бросает вызов самой сущности математики! Основополагающим принципом всего нашего математического наследия и образования является непоколебимая решимость не склоняться перед авторитетом каких-то неясных правил, понять которые мы не надеемся. Мы должны *видеть* —

по крайней мере, в принципе — что каждый этап рассуждений может быть сведен к чему-то простому и очевидному. Математическая истина не есть некая устрашающе сложная догма, обоснованность которой находится вне границ нашего понимания — она строится из подобных простых и очевидных составляющих; и когда они становятся ясны и понятны нам, с их истинностью соглашаются все без исключения.

С моей точки зрения, получить такое явное *reductio ad absurdum* (без применения настоящего математического доказательства) мы даже и мечтать не могли! Основная идея должно быть теперь ясна. Математическая истина — это не то, что мы устанавливаем просто за счет использования алгоритма. Кроме того, я полагаю, что наше *сознание* — это решающая составляющая в нашем понимании математической истины. Мы должны «видеть» истинность математических рассуждений, чтобы убедиться в их обоснованности. Это «вйдение» — самая суть сознания. Оно должно присутствовать *везде*, где мы непосредственно постигаем математическую истину. Когда мы убеждаемся в справедливости теоремы Геделя, мы не только «видим» ее, но еще и устанавливаем неалгоритмичность природы самого процесса «вйдения».

Вдохновение, озарение и оригинальность

Я должен попытаться как-то прокомментировать те внезапные вспышки озарения, которые мы называем вдохновением. Откуда берутся все эти мысли и образы? Может быть, они появляются из нашего бессознательного — или все же сознание существенным образом связано с их рождением? Можно привести множество примеров из воспоминаний великих мыслителей, где они прямо указывали на такие события. Как математик, я особенно интересуюсь теми случаями, когда вдохновение посещало именно математиков, но думаю, что между математикой и другими науками и искусством есть много общего. Эта тема великолепно изложена в небольшой работе «Исследования психологии процессов изобретательства в области математики» — классическом труде выдающегося французского математика Жака Адамара — к которой я и отсылаю читателя. В ней он приводит многочисленные примеры озарения в изложении ведущих математиков и не только. Один из наиболее известных случаев связан с Анри Пуанкаре. В начале Пуанкаре описывает свои напряженные сознательные исследования, связанные с построением так называемых «функций Фукса», которые в конце концов явно зашли в тупик. И вот что он пишет далее:

> «...Я покинул Кон, где я жил в то время, чтобы принять участие в геологической экспедиции, организованной Горной школой. Впечатления от поездки заставили меня забыть о моей математической работе. Достигнув местечка Кутонс мы сели в омнибус, чтобы добраться на нем до следующего пункта назначения. В тот момент, когда я ставил ногу на подножку, мне пришла в голову идея, которая, казалось, никоим образом не вытекала из моих прошлых раздумий, что преобразования, используемые мной для определения функций Фукса, были идентичны определенным преобразованиям в неэвклидовой геометрии. Я не проверил эту идею. У меня просто не было времени, так как когда я занял свое место в омнибусе, я продолжил прерванную беседу — но я был совершенно уверен в правильности моей догадки. Вернувшись в Кон, я выбрал свободное время и, проверив для собственного спокойствия свое предположение, убедился в его справедливости».

Что поражает в этом примере (как и во многих других, приведенных Адамаром) — это внезапность появления столь сложной и глубокой идеи в сознании Пуанкаре, которое в тот момент было занято совершенно другим; и тот факт, что возникновение этой идеи сопровождалось четким ощущением ее истинности, которую полностью подтвердили последующие расчеты. Тут нужно сразу оговориться, что подобные идеи сама по себе далеко не так просты, чтобы их можно было легко выразить словами. Думаю, что для ясного изложения своих мыслей Пуанкаре потребовалось бы провести примерно часовой семинар для экспертов в этой области. Ясно, что эта идея могла полностью

оформиться в сознании Пуанкаре только после долгих часов размышлений, направленных на изучение всех возможных аспектов указанной проблемы. Да, в некотором смысле, идея, осенившая Пуанкаре, когда он садился в омнибус, была «единичной» идеей, которую можно было полностью осознать в один момент. Еще более замечательной представляется убежденность Пуанкаре в ее справедливости — убежденность, которая сделала последующую детальную проверку этой идеи почти что излишней.

Пожалуй, мне стоит попытаться соотнести этот случай с моим собственным опытом, который оказывается в каком-то смысле похожим. На самом деле, я не могу вспомнить ни одной ситуации, когда хорошая идея пришла бы мне в голову «с неба», как это произошло в случае с Пуанкаре (или во многих других известных примерах подлинного вдохновения). Что касается меня, то мне нужно, чтобы выполнялись определенные условия: мне необходимо *думать* о том вопросе, над которым я в данный момент работаю, пусть даже в «фоновом режиме» и не целенаправленно, но обязательно осознанно; вполне возможно, что при этом я буду заниматься чем-то посторонним и успокаивающе-монотонным, например, бриться; вероятно, я заново возьмусь размышлять о проблеме, которая на некоторое время была отложена в сторону. И, конечно же, нужно посвятить не один час упорным сознательным раздумьям, после которых может потребоваться определенное время для того, чтобы вновь переключиться на решаемую проблему. И, тем не менее, ощущение, когда искомое решение возникает при таких условиях подобно «вспышке» — и при этом ты совершенно уверен в его правильности — мне достаточно хорошо знакомо.

Вероятно, стоит привести конкретный пример, который может оказаться небезынтересным и с другой точки зрения. Осенью 1964 года меня занимал вопрос о сингулярностях черных дыр. Оппенгеймер и Снайдер в 1939 году показали, что строго *сферический* коллапс массивной звезды может приводить к образованию центральной сингулярности пространства-времени, выходящей за пределы классической общей теории относительности (см. главу 7, с. 291, 294). Многие считали, что этого неприятного вывода можно было бы избежать, если бы удалось убрать (необоснованное) предположение о *строгой* сферической симметрии. В противном случае получается, что вся коллапсирующая материя стремится к единой центральной точке, где (как вполне закономерно было бы предположить, учитывая симметричность ситуации) возникает сингулярность бесконечной плотности (вещества). Вполне разумным кажется предположение о том, что *без* такой симметрии материя попадала бы в центральную область далеко не так согласованно, вследствие чего сингулярности бесконечной плотности могло бы и не получиться. Возможно даже, что вся материя в этом случае снова «раскрутилась» бы, демонстрируя поведение, совершенно отличное от идеализированной черной дыры Оппенгеймера и Снайдера [3].

Стимулом для моих собственных размышлений на эту тему послужил вновь возникший интерес к проблеме черных дыр, связанный со сравнительно недавно открытыми квазарами (в начале 1960-х годов). Физическая природа этих на удивление ярких (с учетом отделяющих их от Земли расстояний) астрономических объектов вызвала у некоторых специалистов предположение о том, что в центре каждого из них может находиться нечто наподобие черной дыры Оппенгеймера—Снайдера. С другой стороны, многие считали, что гипотеза Оппенгеймера—Снайдера о сферической симметрии может привести здесь к совершенно неверным представлениям. Однако мне пришло в голову (в контексте другой работы, которую я выполнял), что можно было бы сформулировать и доказать точную математическую теорему, указывающую на *неизбежность* возникновения сингулярностей в пространстве-времени (в рамках стандартной общей теории относительности) и тем самым подтверждающую наличие черных дыр — при условии достижения коллапсом определенной «точки необратимости». Я не знал, как математически можно было бы определить такую точку (не используя при этом условия сферической симметрии), не говоря уже о формулировке или доказательстве соответствующей теоремы. В то время приехал

коллега из США Айвор Робинсон, с которым у меня, пока мы шли по улицам Лондона в направлении моего офиса, завязалась оживленная дискуссия на совершенно другую тему. Разговор на момент прекратился, когда мы переходили через дорогу, и был продолжен только на другой ее стороне. И в эти несколько мгновений — я знаю это совершенно точно! — у меня возникла некая идея, которая также быстро оказалась стерта из памяти возобновившейся беседой!

В тот же день, после того, как мой коллега ушел, я вернулся в свой офис. Я помню, что у меня было странное чувство душевного подъема, которое я сам себе никак не мог объяснить. Я начал перебирать все отложившиеся в памяти впечатления этого дня в попытке отыскать причину такого непонятного воодушевления. После исключения множества неподходящих возможностей, я в конце концов вспомнил ту мысль, которая возникла у меня при переходе улицы — и в которой заключалось решение задачи, постоянно крутившейся у меня в голове все последнее время!

Несомненно, это был искомый критерий — который я впоследствии назвал «ловушечная поверхность» — и мне уже не понадобилось много времени, чтобы набросать план доказательства искомой теоремы (Пенроуз [1965]). И хотя прошло еще немало времени, прежде чем было сформулировано математически строгое доказательство — ключевое место в нем сохранила та первоначальная идея, которая пришла мне в голову при пересечении улицы. (Я иногда пытаюсь представить себе, что было бы, если бы в течении этого дня произошло *другое*, менее существенное событие, сравнимое, однако, по эмоциональному воздействию. Может быть, в этом случае, я бы вообще никогда не вспомнил про свою мимолетную идею!)

Эта история подводит меня к еще одному вопросу, связанному с вдохновением и озарением, и касающемся той более чем существенной роли, которую играют при формировании суждений *эстетические* критерии. Можно смело утверждать, что в искусстве эстетические критерии имеют первостепенное значение. Эстетика в искусстве — это сложнейший предмет, изучению которого философ посвящает иной раз всю свою жизнь. Можно было бы утверждать, что в математике, да и в науке вообще, такие критерии скорее второстепенны, а главным критерием всегда является *истинность*. Однако вряд ли возможно отделить одно от другого, когда речь заходит о вдохновении и озарении. У меня создается впечатление, что твердая уверенность в *правильности* идей, приходящих в голову в момент прозрения очень тесно (пусть не полностью коррелируя, но и заведомо не случайно) связана с эстетическими качествами. Красивая идея имеет гораздо больше шансов быть правильной, чем идея нескладная. По крайней мере, об этом свидетельствует мой собственный опыт, а также аналогичные замечания, сделанные другими (см. Чандрасекар [1987]). Вот что, например, пишет Адамар:

> «...ясно, что никакое значительное открытие или изобретение не может быть сделано без *сознательного стремления* к нему. Но в случае с Пуанкаре мы видим и другое — чувство прекрасного, которое сыграло свою роль необходимого *средства* изысканий. И мы приходим к двойному заключению:
>
> что изобретение — это выбор;
> что критерием этого выбора служит чувство научной красоты».

Более того, Дирак [1982], например, непоколебим в убеждении, что именно его *тонкое чувство прекрасного* позволило ему предугадать вид уравнения электрона («уравнение Дирака», упоминаемое на с. 255), в то время как поиски остальных не увенчались успехом.

Я нисколько не сомневаюсь в том, что для меня значение эстетических критериев для мышления трудно переоценить — как в случае ощущения «уверенности» при спонтанном возникновении идей в минуты «вдохновения»; так и в отношении более «прозаических» решений, которые постоянно приходится находить, продвигаясь к желанной цели. Я писал об этом еще в связи с открытием непериодичных «плиточных» наборов замощений, показанных на рис. 10.3 и 4.11. Бесспорно, именно эстетичность первого набора — не только внешний вид, но также и его интригующие математические свойства — позволили мне интуитивно (возможно, в виде «вспышки», но только

лишь с 60 %-ной вероятностью!) понять, что этот узор мог быть создан по определенным правилам состыковки (то есть как мозаика-головоломка). Скоро я собираюсь подробнее рассказать об этих плиточных структурах (см. Пенроуз [1974]).

На мой взгляд очевидно, что эстетические критерии важны не только при формировании спонтанных суждений, являющихся результатом озарения, но и гораздо чаще — в каждом суждении, которое появляется в ходе математической (или, говоря в целом, научной) работы. Строгое доказательство — это обычно *последний* шаг! Перед этим приходится строить множество предположений, и на этом этапе решения, подсказанные эстетическим восприятием, играют исключительно важную роль — конечно, с учетом логически непротиворечивых выводов и известных фактов.

Именно эти суждения я принимаю в качестве критерия сознательного мышления. Я полагаю, что даже при внезапной вспышке озарения, которая, вероятно, является конечным «продуктом» работы бессознательного, арбитром является *сознание*, и идея будет быстро отвергнута и забыта, если, по мнению сознания, она «не звучит». (Забавно, что я *все-таки* забыл о своей «ловушечной поверхности» — но не на том уровне, который я имею в виду. Идея достаточно прочно засела в сознании, чтобы надолго сохранить напоминание о себе.) «Эстетический» запрет, о котором идет речь, мог бы, как я полагаю, вообще закрыть доступ неприглядным идеям к тем уровням сознания, где они могли бы осознанно восприниматься сколь-нибудь длительное время.

Какова моя точка зрения на участие *бессознательного* в рождении «вдохновенной мысли»? Признаюсь, что эти вопросы далеко не так ясны для меня, как хотелось бы. Эта та область, в которой бессознательное, по-видимому, и в самом деле играет крайне важную роль, и я должен согласиться с тем, что бессознательные процессы там весьма существенны. Также приходится признать, что вряд ли бессознательное подбрасывает нам идеи случайным образом. Должен существовать мощнейший механизм отбора, который позволял «тревожить» сознание только тем идеям, у которых «есть шанс». Я готов предположить, что критерии отбора — по большей части носящие своеобразный «эстетический» характер — в значительной степени утверждаются уже с учетом «пожеланий сознания» (подобно ощущению нескладности, возникающему при виде математических идей, которые несовместимы с уже установленными общими принципами).

В связи с этим необходимо затронуть вопрос о том, что представляет собой подлинная *оригинальность*. Мне кажется, что тут действуют два фактора, а именно: процессы «предложения» и «отбора». Из них «предложение» кажется мне по большей части процессом бессознательным, тогда как «отбор» — наоборот. Без эффективного процесса «предложения» новые идеи не возникали бы совсем. Но сама по себе эта процедура мало полезна. Нужен эффективный механизм оценки этих идей, который позволил бы выжить только тем из них, которые представляются достаточно разумными. Во сне, например, необычные идеи возникают легко и в большом количестве — но лишь в очень редких случаях они проходят критический контроль бодрствующего сознания. (Что касается меня, то у меня во сне никогда не возникали плодотворные научные идеи, в то время как другим — например, химику Кекуле при открытии им структуры бензола — кажется, повезло больше.) По моему мнению, именно сознательный процесс «отбора» (или построения суждений) является центральным в содержании понятия оригинальности, а вовсе не бессознательный процесс «предложения»; но я прекрасно понимаю, что многие могут придерживаться противоположной точки зрения.

Прежде чем оставить эту тему в таком неудовлетворительном состоянии, как она есть, я должен упомянуть другую поразительную черту, присущую рожденным в состоянии вдохновения идеям, а именно — их *масштабность*. История Пуанкаре, рассказанная выше, являет собой поразительный пример проявления этого свойства, поскольку идея, мимолетно возникшая в его голове, должна была охватывать весьма обширную область математической мысли. Возможно, более наглядным для читателя-нематематика (хотя и столь же

непостижимым) является способ, которым (иные) художники могут представить себе весь замысел своего творения целиком. Подобный удивительный случай очень живо описан Моцартом (см. Адамар [1945], с. 16):

> «Когда я чувствую себя хорошо и нахожусь в добром расположении духа; или когда я предпринимаю поездку или отправляюсь на прогулку после сытной трапезы; или ночью во время бессонницы — мне в голову приходит сколько угодно самых разных идей. Откуда и как они приходят? Я не знаю и ничего не могу с этим поделать. Те, которые мне приятны, я удерживаю в голове и часто напеваю без слов; так, по крайней мере, мне говорили. Когда у меня возникает тема, сразу же приходит следующая мелодия, соединяясь с первой согласно требованиям композиции в целом: контрапункт, партия каждого инструмента — и, наконец, все музыкальные фрагменты складываются в завершенное произведение. Тогда моя душа горит вдохновением. Произведение растет; я постоянно дополняю его, прорабатываю все более мелкие детали, пока в один прекрасный момент композиция не оказывается полностью сформирована у меня в голове, хотя она может быть и довольно длинной. Тогда мой ум охватывает ее единым взглядом, как красивую картину или прекрасную девушку. Это не последовательный процесс, при котором различные части произведения прорабатываются до мелочей и стыкуются друг с другом (так, как это будет сделано в дальнейшем) — нет, я слышу его целиком, как это позволяет мое воображение».

Мне кажется, что это хорошо согласуется с моей схемой «предложения-отбора». «Предложение» представляется здесь бессознательным («я ничего не могу с этим поделать»), хотя и, несомненно, в высокой степени избирательным, в то время как «отбор» выполняет функцию сознательного судьи вкуса («те, которые мне приятны, я удерживаю...»). Масштабность идеи, рожденной в минуты вдохновения, особенно отчетливо проглядывает в высказывании Моцарта («это не последовательный процесс... нет, я слышу его (произведение) целиком») и Пуанкаре («Я не проверил эту идею. У меня просто не было времени...»). Более того, я готов настаивать, что нашему сознательному мышлению в целом подобная масштабность присуща изначально. К этому вопросу я еще вернусь.

Невербальность мысли

Одно из главных утверждений, которое Адамар делает в своей работе о творческом мышлении — это убедительное опровержение популярного сегодня тезиса, который гласит, что вербализация необходима для формирования мысли. Вряд ли здесь можно привести возражение более убедительное, чем то, которое содержится в письме Альберта Эйнштейна Адамару:

> «Слова или язык, как в устной, так и в письменной форме, по-видимому, не играют никакой роли в механизме моего мышления. Психические сущности, которые, по-видимому, и являются составляющими элементами мысли — это определенные знаки и более или менее отчетливые образы, которые могут „произвольно" воспроизводиться и комбинироваться по собственному желанию... В моем случае, упомянутые элементы носят визуальный и моторный характер. Общепринятые слова или другие знаки мне приходится подбирать только на второй стадии, когда упомянутые ассоциативные связи приобретают отчетливые очертания и могут быть воспроизведены по моей воле».

Еще здесь стоит процитировать видного генетика Фрэнсиса Гальтона:

> «Для меня серьезную трудность представляет письмо, а еще бóльшую — словесное изъяснение, так как размышления в словесной форме даются мне далеко не так легко, как в любой другой. Часто случается, что проделав большую работу и получив результаты, которые мне абсолютно ясны и вполне меня удовлетворяют, при попытке выразить их словами я сталкиваюсь с необходимостью переводить себя в совершенно иную интеллектуальную плоскость. Мне приходится перекладывать свои мысли на язык, который не слишком-то хорошо им соответствует. Поэтому я вынужден тратить уйму

времени в поисках подходящих слов и фраз, и часто осознаю, что, выступая без подготовки, бываю не понят не из-за неясности содержания высказывания, а только лишь из-за неуклюжести своих вербальных конструкций. Это один из небольших, но досадных моих недостатков».

Нечто сходное пишет и сам Адамар:

«Я утверждаю, что слова полностью отсутствуют в моей голове, когда я действительно *предаюсь* раздумьям, и я нахожу случай Гальтона полностью идентичным моему личному опыту, поскольку и у меня самого даже после прочтения или выслушивания вопроса все слова исчезают в тот самый момент, когда я начинаю их обдумывать; и я полностью согласен с Шопенгауэром, когда он пишет: „Мысли умирают в момент, когда воплощаются в слова“».

Я цитирую эти примеры, потому что они очень хорошо согласуются с моим собственным способом мышления. Почти все мое математическое мышление визуализируется или протекает на уровне не-вербальных понятий, хотя мысли очень часто сопровождаются пустыми и почти бесполезными словесными комментариями, такими как «вот это идет с этим, а это — с этим». (Иногда я могу употреблять слова для выражения простых логических выводов.) Трудности, которые испытывали упомянутые ученые при переводе своих мыслей на язык слов, я часто испытывал и сам. Причиной тому в большинстве случаев служило просто-напросто отсутствие адекватных терминов, способных выразить требуемые понятия. Действительно, я часто веду расчеты, используя специально разработанные диаграммы, которые представляют собой сокращенную запись определенных типов алгебраических выражений (см. Пенроуз и Риндлер [1984]). Необходимость перевода таких диаграмм в слова — это очень трудоемкий процесс, и я это делаю только в случае крайней необходимости, когда нужно подробно объяснить что-то другим. И еще одно наблюдение: я случайно заметил, что если сосредотачиваю все свое внимание на математике и некоторое время занимаюсь только ей, а потом кто-то внезапно обращается ко мне, то в течение нескольких следующих секунд я почти не способен говорить.

Не могу сказать, что я никогда не думаю в словесной форме — просто я нахожу слова почти бесполезными для *математического* мышления. Другие виды рассуждений, возможно, такие, как *философские*, являются, вероятно, гораздо более подходящими для вербального выражения. Может быть, поэтому так много философов считают язык неотъемлемым средством интеллектуальной деятельности и сознательного мышления! Нет сомнения, что каждый человек думает по-своему — это подтверждает и мой собственный опыт, и мнения других математиков. Наиболее полярными стилями математического мышления являются, как кажется, аналитический/геометрический. Интересно, что Адамар считал себя аналитиком, хотя использовал скорее визуальные, чем вербальные образы в своем математическом мышлении. Что касается меня, то я в значительной степени тяготею к геометрическим методам. Если же говорить обо всех математиках, то разброс здесь окажется весьма широк.

Но коль скоро мы согласились с тем, что значительная часть сознательного мышления, на самом деле, может иметь невербальный характер — а с моей точки зрения к другому выводу приведенные выше соображения привести не могут — тогда, наверное, читателю будет нетрудно поверить также и в то, что подобное мышление может иметь неалгоритмическую составляющую!

Напомню, что в главе 9 (с. 330) я упоминал о часто встречающейся точке зрения, согласно которой только одно полушарие мозга — то, где находится центр речи (левое у большинства людей) — способно также и на сознательное мышление.

После ознакомления с вышеизложенным читателю должно быть ясно, почему я считаю эту точку зрения совершенно неприемлемой. Я не знаю, используют ли, как правило, математики одно полушарие чаще, чем другое; но нет сомнения в том, что для истинного математического мышления необходим высокий уровень сознания. В то время как аналитическое мышление, по всей видимости, сосредоточено в левой

половине мозга, геометрическое мышление, напротив, часто приписывают *правой* половине; так что вполне разумной является предположение о том, что значительная часть *сознательных* математических рассуждений проводится *все-таки* в правом полушарии!

Сознание у животных?

Прежде чем закончить рассуждения о важности вербализации применительно к сознанию, я должен еще рассмотреть один вопрос, который вкратце уже затрагивался ранее, а именно — могут ли живые существа, отличные от нас, обладать сознанием? Мне кажется, что люди иногда используют неспособность животных говорить как аргумент, отрицающий саму возможность наличия у них достаточно развитого сознания и, как следствие, позволяющий априори отказывать им в самых элементарных «правах». Читатель может легко догадаться, что для меня подобные рассуждения являются неубедительными, поскольку для многих сложных разновидностей сознательного мышления (например, для математического) вербализация и вовсе не требуется. А некоторые по той же причине — т. е. из-за отсутствия речевых способностей — считают, например, что правая сторона мозга обладает сознанием «не более», чем шимпанзе (см. Леду [1985], с. 197–216).

Существуют значительные разногласия относительно способности горилл и шимпанзе выражать свои мысли при помощи *языка знаков* (а не обычной человеческой речи, которую они не могут воспроизводить из-за особенностей строения их голосовых связок; см. статьи Колина Блэйкмора и Сьюзан Гринфилд (Блэйкмор, Гринфилд [1987])). Хотя полемика еще продолжается, один факт уже не вызывает сомнений: эти человекообразные обезьяны могут общаться таким образом, по крайней мере, на некотором элементарном уровне. По моему глубокому убеждению, отказываться признать в этом общении «вербализацию» — это прямое проявление высокомерия со стороны тех, кто придерживается такой точки зрения. Видимо, закрывая обезьянам доступ в «клуб способных к вербализации», надеются автоматически исключить их и из «клуба обладающих сознанием»!

Оставляя пока в стороне вопрос о речи, обратимся к убедительным свидетельствам, которые указывают на способность шимпанзе к подлинному «вдохновению». Конрад Лоренц описывает шимпанзе в комнате, где к потолку был подвешен банан, до которого обезьяна не могла достать, а в одном из углов был поставлен ящик:

> «Задача не давала ему покоя, и он возвращался к ней вновь и вновь. Затем внезапно — по-другому и не скажешь — его прежде унылая физиономия „озарилась“. Взгляд шимпанзе перемещался с банана на пустое пространство под ним, оттуда на ящик, потом снова на место под бананом, и оттуда на банан. В следующий момент он издал крик радости и кувыркнулся в сторону ящика, явно пребывая в превосходнейшем настроении. Совершенно уверенный в успехе, он толкнул ящик под банан. Могу поспорить, что никто из видевших его в тот момент не усомнился бы в способности человекообразных обезьян к таким прозрениям, испытав которые, впору воскликнуть „Эврика!“».

Обратите внимание, что точно так же, как в случае с Пуанкаре, когда тот садился в омнибус, шимпанзе был «совершенно уверен в успехе» еще до того, как он проверил свою идею. И если я прав, утверждая, что подобные суждения требуют участия сознания, то перед нами оказывается неопровержимое свидетельство того, что животные действительно могут обладать сознанием.

Глядя на дельфинов (и китов), мы невольно задаемся одним интригующим вопросом. Как нетрудно заметить, головной мозг дельфинов имеет такие же (или даже бóльшие) размеры, как и наш собственный; а кроме того, дельфины могут посылать друг другу чрезвычайно сложные звуковые сигналы. Вполне возможно, что такой большой мозг нужен для каких-то иных целей, которые не сводятся к «интеллектуальной» деятельности в человеческом или околочеловеческом понимании. Более того: не имея рук, приспособленных для хватания, они не могут создать «цивилизацию», которую

мы были бы способны оценить. И хотя они по той же самой причине не могут писать книг, они вполне способны время от времени превращаться в философов и размышлять о смысле своей жизни! Что, если они иногда передают свое ощущение «самосознания» при помощи этих сложных звуковых сигналов, распространяющихся под водой? Я не встречал ни одного исследования, где бы изучалось, используют ли дельфины какую-то одну определенную сторону мозга для «вербализации» и общения друг с другом. В связи с проведенными на людях операциями по разделению мозга, которые загадочным образом влияли на целостность «я» человека, следует отметить еще одну особенность дельфинов: их полушария никогда не погружаются в сон [4] одновременно — вместо этого каждая сторона мозга спит по очереди. Согласитесь, хорошо было бы выяснить у них, как *они* «ощущают» целостность своего сознания!

Соприкосновение с миром Платона

Я уже упоминал о том, что разные люди скорее всего мыслят по-разному — и даже у разных математиков мысли при решении математической задачи формируются не одинаково. Я вспоминаю, что, поступая на математический факультет университета, я ожидал, что мои будущие коллеги-математики должны думать примерно так же, как я. В школе мои одноклассники, казалось, думали совсем иначе, чем я, что меня несколько удручало. «Теперь, — думал я с восторгом, — я найду коллег, с которыми общаться мне будет гораздо легче! Некоторые будут мыслить более продуктивно, чем я, а некоторые — менее; но все они смогут настроиться на мою ментальную длину волны!» Как же я заблуждался! Думаю, что тогда я познакомился с гораздо бо́льшим числом различных способов мышления, чем за все предыдущее время! Да, мой собственный образ мыслей был куда более геометрическим и далеко не столь аналитическим по сравнению с остальными — но у них было и множество других различий в способе мышления. У меня всегда вызывало затруднение понимание словесного описания формулы, в то время как у многих из моих коллег, казалось, с этим не возникало никаких трудностей.

Довольно часто случалось так, что, слушая своего коллегу, пытающегося объяснить мне какую-нибудь математическую выкладку, я практически совсем не улавливал логической связи между следующими друг за другом наборами слов. Однако, в моей голове постепенно формировалась догадка о содержании передаваемых мне идей — причем складывалась она в рамках моей собственной терминологии и, скорее всего, была мало связана с ментальными образами, которыми оперировал мой коллега, обращаясь к данной проблеме, — и тогда я отвечал. К моему удивлению, эти ответы чаще всего воспринимались как адекватные, и беседа продолжала развиваться в таком же ключе, причем к концу становилось ясно, что состоялся поистине позитивный обмен мнениями. Однако сами предложения, которые произносил каждый из нас в ходе беседы, чаще всего оставались не поняты! В последующие годы, будучи уже профессиональным математиком (или физиком-математиком), я пришел к выводу, что ситуация в целом практически не изменилась по сравнению с тем временем, когда я учился на младших курсах. Возможно, с увеличением моего математического багажа я стал несколько лучше разбираться, о чем говорят другие, пытаясь донести до моего сознания определенную мысль; и, наверное, я научился адаптировать свой стиль изложения, каждый раз подстраиваясь под конкретного слушателя. Однако, в сущности, все осталось по-прежнему.

Для меня часто является загадкой, как вообще возможно *подобное* общение, но теперь я все же осмелюсь дать некоторое объяснение, которое, как мне кажется, могло бы иметь самое непосредственное отношение к уже затронутым ранее вопросам. Суть здесь заключается в том, что при общении математиков происходит *не* только обмен *фактами*. Чтобы состоялась передача ряда фактов от одного собеседника другому, первому из них необходимо излагать эти факты достаточно понятно, а второму — воспринять каждый из них в отдельности. Но в математике *фактическое* содержание играет второстепенную роль. Математические утвержде-

ния являются с необходимостью истинными (или же с необходимостью ложными!), и даже если первый математик своим утверждением только нащупывает искомую истину, то именно эту истину воспримет его собеседник (конечно, если исходное утверждение будет им правильно понято). Ментальные конструкции второго математика могут в деталях отличаться от тех образов, которые возникают у первого, равно как могут отличаться и их словесные описания — но соответствующая математическая идея в результате все-таки будет передана.

Такой тип общения был бы совершенно невозможен, если бы не то обстоятельство, что *интересные* или *глубокие* математические истины растворены (с небольшой плотностью) в массе всех возможных математических истин. Если бы передаваемая истина заключала в себе, скажем, *неинтересное* утверждение наподобие $4897 \times 512 = 2\,507\,264$, то второму собеседнику, естественно, придется полностью понять первого, иначе это точное утверждение не сможет быть передано. Но при сообщении математически *интересного* утверждения часто удается понять его интуитивно, даже если для его описания использовались расплывчатые образы и понятия.

Это может показаться парадоксальным, поскольку математика — это предмет, где точность всегда ставится превыше всего. В самом деле, в письменных отчетах большое внимание уделяется точной формулировке и завершенности всех утверждений. Однако, чтобы передать математическую идею (обычно посредством словесного описания), такая точность иной раз является помехой, так что вначале может потребоваться менее четкая описательная форма. А как только будет понята самая суть идеи — тогда можно уже переходить и к деталям.

Как же получается, что математические идеи могут передаваться подобным образом? Лично мне представляется, что всякий раз, когда ум постигает математическую идею, он вступает в контакт с миром математических понятий Платона. (Вспомним, что, по Платону, математические идеи имеют собственное бытие и населяют некий идеальный мир, доступ в который осуществляется только благодаря работе интеллекта (см. с. 108, 154).) Когда человек «видит» математическую истину, его сознание пробивается в этот мир идей и устанавливает с ним кратковременный прямой контакт (т. е. осуществляет «доступ посредством интеллекта»). Я описал это «вйдение» в связи с теоремой Геделя, хотя, вообще говоря, здесь заключена сущность математического понимания. Общение математиков становится возможным постольку, поскольку у каждого из них в этот момент есть *прямой путь к истине*, а сознание каждого способно при этом постигать математические истины непосредственно, путем «вйдения». (В самом деле, часто акт понимания сопровождается словами типа «О, я вижу!»[4].) Так как каждый математик может установить непосредственный контакт с миром идей Платона, то общение их друг с другом проходит значительно легче, чем это можно было бы ожидать. Ментальные образы, возникающие у каждого из них, когда осуществляется соприкасание с миром Платона, могут быть существенно различными, но общение тем не менее возможно, поскольку каждый находится в прямом контакте с *одним и тем же* существующим вне нас миром Платона!

В соответствии с этой точкой зрения, наш ум всегда способен на подобный прямой контакт. Но за один раз можно продвинуться лишь на немного. Математическое открытие как раз и состоит в расширении области контакта. Поскольку математические истины являются с необходимостью истинами, никакой содержательной «информации» в общепринятом смысле этого слова исследователь не получает. Вся информация уже находилась там изначально. Все, что требовалось — это соединить разные части друг с другом и «увидеть» ответ! Это очень хорошо согласуется с представлениями самого Платона о том, что (скажем, математическое) открытие — это всего лишь одна из форм *воспоминания*! В самом деле, меня часто поражало сходство между двумя состояниями, когда ты мучительно стараешься вспомнить чье-то имя — и когда пытаешься найти адекватное математическое

---

[4] В английском языке фраза *Oh, I see!* («О, я вижу!») по смыслу эквивалентна возгласу «О, я понимаю!». — *Прим. ред.*

понятие. В обоих случаях искомое в некотором смысле *уже присутствует* в голове, хотя во втором случае «вспоминание» математической идеи связано с необычной формой вербализации.

Чтобы такие идеи были полезны для объяснения принципов математического общения, нужно представить себе, что интересные и глубокие математические идеи отличаются способностью к более «основательному» существованию, чем неинтересные или тривиальные. Эти соображения пригодятся нам при рассмотрении ряда умозрительных заключений, приведенных в следующем разделе.

Взгляд на физическую реальность

Любой подход к вопросу о возникновении сознания в царстве физической реальности неявно подразумевает необходимость определения природы самой физической реальности.

Концепция «сильного» ИИ, например, заключается в том, что «разум» возникает через воплощение достаточно сложного алгоритма, по мере того как этот алгоритм реализуется различными объектами физического мира. При этом, сущность этих объектов, согласно данной теории, значения не имеет. Нервные сигналы, электрический ток, идущий по проводам, винтики, приводные ремни или водопроводные трубы — все это одинаково подходит в качестве «аппаратной части». Алгоритм рассматривается как нечто самодостаточное. Однако, для предоставления алгоритму возможности «существовать» вне зависимости от любой физической реализации совершенно необходимо разделять воззрения Платона на природу математики. Сторонникам «сильного» ИИ было бы нелегко согласиться с альтернативной точкой зрения о том, что «математические понятия существуют только лишь в умах», поскольку при этом образуется замкнутый круг, где для возникновения умов требуется наличие алгоритмов, а для изначального существования алгоритмов — наличие умов! Они могли бы попытаться придерживаться иного мнения, согласно которому алгоритмы могут существовать как знаки на листе бумаги, или как магнитные силовые линии в кусках железа, или как смещения зарядов в памяти компьютера. Но такое упорядочение материала само по себе не есть алгоритм. Чтобы стать алгоритмом, оно должно иметь *интерпретацию*, т. е. должна существовать возможность *декодирования* этого порядка; а это будет зависеть от «языка», на котором написан алгоритм. Снова оказывается, что нужен заранее существующий разум, чтобы «понимать» этот язык, и мы возвращаемся к исходной точке. Предположив затем, что алгоритмы населяют мир Платона, и именно там обитает разум (согласно теории «сильного» ИИ), мы оказываемся перед вопросом отношения между физическим миром и миром Платона. Что, на мой взгляд, является в теории «сильного» ИИ аналогом проблемы «ум—тело»!

Я же придерживаюсь иной точки зрения, поскольку считаю, что (сознательный) ум существенно *неалгоритмичен*. Но меня несколько смущает то, что между взглядами сторонников «сильного» ИИ и моими собственными существует большое число точек соприкосновения. Я уже указывал, что считаю сознание тесно связанным со способностью воспринимать несомненные истины, и тем самым осуществлять прямой контакт с миром математических понятий Платона. Это неалгоритмическая процедура — и отнюдь не алгоритмы должны населять тот мир, который так важен для нас — и вновь проблема «ум—тело», если следовать этой точке зрения, оказывается тесно связанной с вопросом об отношении мира Платона к «реальному» миру существующих физических объектов.

В главах 5 и 6 мы видели, насколько хорошо реальный физический мир согласуется с некоторыми исключительно точными математическими теориями (ПРЕВОСХОДНЫЕ теории, см. с. 149). И эта поразительная точность неоднократно подчеркивалась многими исследователями (см., в частности, Вигнер [1960]). Мне трудно поверить — хотя многие считают это неоспоримым — в справедливость утверждения о том, что ПРЕВОСХОДНЫЕ теории могли возникнуть просто в результате случайного естественного отбора идей, при котором «выживают» только лучшие. В это трудно

поверить потому, что уж *слишком* они хороши для того, чтобы оказаться среди идей, возникших случайным образом. На самом деле должна быть какая-то фундаментальная взаимосвязь между математикой и физикой, т. е. между миром Платона и физическим миром.

Говоря о «мире Платона», мы приписываем ему некоторый вид реальности, которая определенным образом сравнима с реальностью физического мира. С другой стороны, сама реальность физического мира кажется уже менее очевидной, чем она представлялась до появления теорий относительности и квантовой механики (относящихся к ПРЕВОСХОДНЫМ теориям) (см. с. 149, 150 и, в особенности, 252). Сама точность этих теорий обеспечивает почти математический абстрактный уровень существованию нашей физической реальности. Не является ли это своего рода парадоксом? Как может конкретная реальность превратиться в абстрактную, да еще и математическую? Возможно, это оборотная сторона вопроса о том, как абстрактные математические понятия могут становиться почти ощутимо реальными в мире Платона. Возможно, в каком-то смысле, эти два мира, на самом деле — *один и тот же* мир? (См. Вигнер [1960], Пенроуз [1979a], Барроу [1988], а также Эткинс [1987].)

Хотя я с большой симпатией отношусь к идее такого отождествления двух миров, вопрос этим далеко не исчерпывается. Как я уже упоминал в главе 3, и выше в этой главе, некоторые математические истины в мире Платона кажутся «более реальными» («более глубокими», «более интересными», «более многообещающими»?), чем остальные. Такими должны быть все те истины, которые наиболее тесно связаны с явлениями физической реальности. (Примером здесь может служить система комплексных чисел (см. главу 3), которая является основной составляющей квантовой механики, ибо амплитуды вероятности выражаются через комплексные числа.) Если принимать в расчет возможность такого отождествления, то становится более понятным, как «разум» мог бы играть роль таинственного связующего звена между физическим миром и математическим миром Платона. Вспомним также (см. главу 4), что есть много областей математического мира — более того, наиболее глубоких и интересных его областей — которым присущ неалгоритмический характер. Поэтому было бы разумным, основываясь на той точке зрения, которую я здесь излагаю, предположить, что неалгоритмические процессы должны играть в физическом мире весьма существенную роль. И я склоняюсь к тому, что эта роль тесно связана с самим понятием «разума».

Детерминизм и жесткий детерминизм

До сих пор было мало сказано о вопросе «свободы воли», который обычно считается неотъемлемым при рассмотрении *активной* составляющей проблемы «ум—тело». Вместо этого, я уделил основное внимание предположению о наличии существенно *неалгоритмической* составляющей в той роли, которую играет осознанное действие. Обычно тема свободы воли обсуждается в связи с детерминизмом в физике. Вспомним, что в большинстве существующих ПРЕВОСХОДНЫХ теорий типа существует явно выраженный детерминизм: если известно состояние системы в определенный момент времени [5], то оно полностью определяется в любой более поздний (или ранний) момент из уравнений теории. Таким образом, по-видимому, для «свободы воли» не остается места, поскольку будущее поведение системы кажется полностью обусловленным физическими законами. Даже U-часть квантовой механики имеет такой же полностью детерминистский характер. Однако R-часть, связанная с «квантовым скачком», не является детерминистской, внося элемент случайности в эволюцию системы во времени. Был момент, когда исследователи старались найти именно здесь свободу воли, полагая, что действие сознания может непосредственно влиять на «скачок» отдельной квантовой системы. Но если R-часть *действительно* случайна, то это тоже нам не слишком поможет, если мы хотим конструктивное применение нашей свободе воле.

Моя собственная точка зрения (правда, не очень четко сформулированная в этом

случае) заключается в том, что должен быть применен некий новый подход (ПКТГ; см. главу 8), который работал бы на границе между квантовой и классической физикой, интерполируя между **U** и **R** (каждая из которых теперь рассматривается как аппроксимация); и этот подход должен содержать *существенно неалгоритмический* элемент. А это подразумевает, что будущее *не будет вычислимым* на основе настоящего, даже если оно им и *определяется*. Я пытался по возможности наиболее ясно определить смысловые различия терминов «вычислимость» и «детерминизм» в главе 5. Мне кажется, что ПКТГ может быть детерминистской, но невычислимой теорией[5]. (Вспомним невычислимую «игрушечную модель», которую я описал в главе 5, с. 164.)

Многие при этом считают, что даже классический (или U-квантовый) детерминизм не является детерминизмом *в полном* смысле этого слова, поскольку исходные условия в принципе не могут быть известны с такой точностью, которая действительно позволила бы просчитать будущее. Иногда совсем небольшие изменения исходных условий могут привести к очень значительным различиям в конечном результате. Именно так возникает «хаос» в (классической) детерминистской системе — явление, приводящее, например, к неопределенностям в прогнозе погоды. Однако очень трудно поверить, что этот вид классической неопределенности может позволить нам сохранять (иллюзорную?) веру в существование свободы воли. Будущее поведение все равно будет *детерминированным* в каждый момент времени, начиная с Большого взрыва, даже если мы окажемся не в состоянии его вычислить (см. с. 167).

То же самое возражение может быть выдвинуто и против моей идеи о том, что *невычислимость* связана скорее с особенностями законов динамики — которые в этом случае считаются исходно неалгоритмическими — чем с нехваткой информации о начальных условиях. *Невычислимое* будущее, согласно этой точке зрения, все равно будет полностью *обусловлено* прошлым — вплоть до момента Большого взрыва. На самом деле я не настолько привержен догмам, чтобы настаивать на том, что методы ПКТГ должны быть по сути детерминистскими, но невычислимыми. Я полагаю, что искомая теория должна иметь более тонкий характер, вследствие чего подобное грубое описание будет к ней просто неприменимо. Единственное, на чем я настаиваю — так это на необходимости присутствия в ней существенно неалгоритмических элементов.

Завершая этот раздел, я хотел бы упомянуть еще об одном представлении о природе детерминизма, причем из числа весьма радикальных. Я называю его *жестким детерминизмом* (Пенроуз [1987б]). Согласно этой теории, не просто будущее предопределяется прошлым — *вся история вселенной* оказывается *раз и навсегда* определенной в соответствии с некоторой точной математической схемой. Такая концепция могла бы привлечь тех, кто склонен каким-нибудь образом отождествлять мир Платона с физическим миром — ибо застывший навеки мир Платона с его однозначной определенностью не оставляет в этом случае вселенной никаких «альтернативных возможностей»! (Я иногда задаю себе вопрос: мог ли Эйнштейн иметь в виду подобную схему, когда он писал: «Что меня собственно интересует, это следующее: мог ли Бог сотворить мир другим, оставляет ли какую-то свободу требование логической простоты» (письмо Эрнсту Штрауссу; см. Кузнецов [1980], с. 363).)

С одним из вариантов жесткого детерминизма мы сталкиваемся в квантово-механической концепции *«множественности миров»* (см. главу 6, с. 260). В соответствии с ней, вышеупомянутая точная математическая схема определяла бы не *единственную* отдельную историю вселенной, а всю совокупность из мириадов мириадов «возможных» историй вселенной. Несмотря на малопривлекательный характер (по крайней мере для меня) такой схемы и множество проблем и несоответствий, которые она в себе несет, мы все же не имеем права сбрасывать ее со счетов как потенциально возможную.

Мне кажется, что, если принять жесткий детерминизм, но *без* множественности миров, то математическая схема, ко-

[5] Стоит отметить, что существует по меньшей мере один подход к квантовой теории гравитации, который, по-видимому, включает элемент невычислимости (Герох, Хартли [1986]).

торая управляет структурой вселенной, вероятно, *должна* быть неалгоритмической [6]. Ибо в противном случае можно было бы, в принципе, просчитать свои будущие действия, а затем вдруг «решить» сделать нечто совершенно другое — получаем очевидное противоречие между «свободой воли» и жестким детерминизмом нашей теории. Вводя не-вычислимость, можно избежать этого противоречия, хотя должен признаться, что я не вполне уверен в адекватности решения такого типа и предвижу в будущем гораздо более тонкое описание «реально действующих» (неалгоритмических) правил, которым подчиняется наш мир!

Антропный принцип

Насколько важно сознание для вселенной в целом? Могла бы вообще вселенная существовать без населяющих ее сознательных существ? Намеренно ли законы физики задумывались такими, чтобы обеспечить существование сознательной жизни? Является ли наше место во вселенной — как в пространстве, так и во времени — каким-то особенным? Вот вопросы, которые ставит перед нами научная гипотеза, известная как *«антропный принцип»*.

Этот принцип может принимать различные формы (см. Барроу, Типлер [1986]). Наиболее приемлемая из них затрагивает только вопрос пространственно-временно́го расположения сознательной (или «разумной») жизни во вселенной. Это — *«мягкий»* антропный принцип. Он может использоваться для объяснения того, почему условия оказались именно такими, что в современную эпоху стала возможна жизнь на Земле. Ответ прост: ведь если бы не было подходящих условий, то мы должны были бы находиться где-то в другом месте, и в иное (благоприятное) время. Этот принцип был очень эффективно использован Брэндоном Картером и Робертом Диком, чтобы разрешить вопрос, остававшийся для физиков загадкой на протяжение многих лет. Вопрос касается существования определенных числовых соотношений между физическими константами (гравитационная постоянная, масса протона, возраст вселенной и т. д.). Интригующим в этих закономерностях был тот факт, что некоторые из них сложились только в настоящую эпоху истории Земли, тем самым указывая на некую — быть может, случайную — исключительность нашего положения во времени (с точностью до нескольких миллионов лет, разумеется!). Впоследствии Картер и Дик нашли этому следующее объяснение: предположили, что эта эпоха совпадает с временем жизни так называемых звезд главной последовательности, одной из которых является наше Солнце. В любую другую эпоху, согласно их утверждениям, нигде и близко не было бы разумной жизни, чтобы измерить те самые физические константы — так что совпадение *должно* было иметь место просто потому, что разумная жизнь возникла бы только в то время, когда *есть* подобное совпадение!

«Жесткий» антропоцентрический принцип идет еще дальше. В этом случае мы рассматриваем наше уникальное положение в пространстве-времени не только этой вселенной, но и бесконечного множества других *возможных* вселенных. Исходя их этого, мы можем сделать ряд предположений относительно того, почему физические константы — или, в более широком смысле, законы физики — как будто специально были спроектированы так, чтобы разумная жизнь вообще могла существовать. Допустим, что константы (или законы) отличались бы от наблюдаемых — тогда мы просто *не могли бы* появиться в этой вселенной и *должны были бы* оказаться в некоторой другой! По моему мнению, достоинства «жесткого» антропоцентрического принципа несколько сомнительны, и теоретики прибегают к нему всякий раз, когда не находят адекватной теории для объяснения наблюдаемых фактов (в первую очередь, это касается теорий физики частиц, где за отсутствием разумного объяснения массам частиц, предполагается, что если бы их значения отличались от настоящих, то жизнь, вероятнее всего, была бы вообще невозможна, и т. д.). С другой стороны, «мягкий» антропный принцип представляется мне безупречным при условии, что им пользуются крайне осмотрительно.

Взяв на вооружение антропный принцип — либо в «жесткой», либо в «мягкой» формах, — можно попытаться показать, что

зарождение сознания было *неизбежно* благодаря тому факту, что сознательные существа, то есть «мы», должны были присутствовать, чтобы наблюдать этот мир — так что *нет* необходимости предполагать, как это делал я, будто способность осознавать даст какое-то преимущество в процессе естественного отбора! По моему мнению, этот довод технически корректен, и доказательство, опирающееся на «мягкий» антропный принцип (по крайней мере), *могло бы* указать на причину, по которой сознание существует в нашем мире независимо от благоволения к нему естественного отбора. С другой стороны, я не могу поверить в то, что антропный принцип и есть та *настоящая* (или единственная) причина, которая обеспечивает эволюцию сознания. Существует достаточно много самых разнообразных свидетельств, способных утвердить меня во мнении, что сознание на самом деле *является* сильным преимуществом в процессе естественного отбора, и что, следовательно, совсем необязательно апеллировать к антропному принципу.

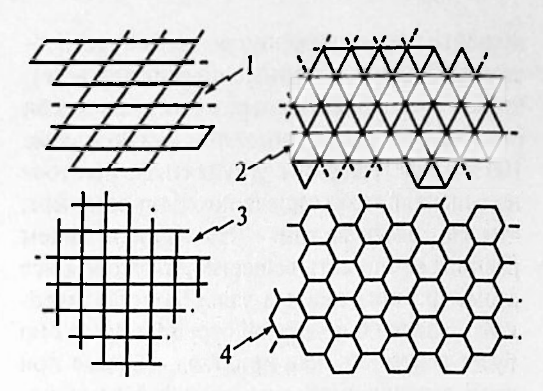


Рис. 10.2. Периодические плиточные замощения с разными типами симметрии (где в каждом случае центр симметрии совпадает с центром плитки): 1) с осью второго порядка; 2) с осью третьего порядка; 3) с осью четвертого порядка; 4) с осью шестого порядка

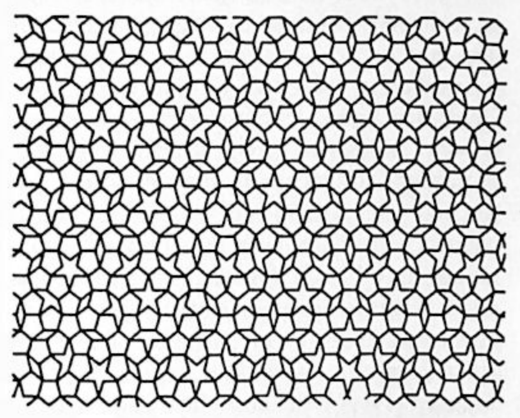


Рис. 10.3. Квазипериодическая плиточная структура (следует заметить, что она образована посредством объединения образцов с рис. 4.11) с кристаллографически «невозможной» квазисимметрией с осью пятого порядка

«Плиточные» структуры и квазикристаллы

Теперь я отойду от масштабных обсуждений последних нескольких разделов и сосредоточусь на обсуждении вопросов, которые, хотя и являются до некоторой степени дискуссионными, все же гораздо более научны и «осязаемы». Возможно, вначале эти рассуждения покажутся отклонением от темы, однако, их важность для нас станет очевидной уже в следующем разделе.

Вспомним примеры «плиточных» замощений, изображенные на рис. 4.12 (с. 138). Эти образцы интересны потому, что они «почти» нарушают общепринятую математическую теорему о кристаллических решетках, которая утверждает, что для кристаллических решеток возможны только симметрии с осью второго, третьего, четвертого и шестого порядков. Под кристаллической решеткой я подразумеваю дискретную систему точек, которая обладает *трансляционной симметрией*. Это означает, что можно определенным образом перемещать решетку без вращения так, чтобы она переходила сама в себя (иными словами, в результате такого сдвига она не изменяется) — а, значит, у такой решетки будет существовать *параллелограмм периодов* (см. рис. 4.8). Примеры «плиточных» замощений с этими разрешенными теорией типами вращательной симметрии показаны на рис. 10.2. С другой стороны, покрытия на рис. 4.12, как и изображенные на рис. 10.3 (которые, в сущности, представляют собой замощения, образованные соединением решеток, изображенных на рис. 4.11, с. 138), *почти* имеют трансляционную симметрию и *почти* обладают симметрией вращения с осью пятого порядка, где «почти» означает, что мож-

но найти такие движения решеток (соответственно, трансляционные и вращательные), при которых решетка переходит сама в себя с любой наперед заданной точностью (кроме 100 %-ной). Не стоит углубляться, что точно означает это утверждение. Единственное, что нам здесь важно — это если в нашем распоряжении есть вещество, в котором все атомы расположены в узлах кристаллической решетки с подобной структурой, то оно будет выглядеть, как кристалл, обладая при этом запрещенной симметрией с осью пятого порядка!

В декабре 1984 году израильский физик Дэни Шехтман, работавший вместе с коллегами в Национальном бюро стандартов в США, в Вашингтоне, объявил об открытии фазы алюминиево-марганцевого сплава, который был похож на кристаллоподобное вещество — теперь называемое *квазикристаллом* — с осью пятого порядка. На самом деле, у этого квазикристаллического вещества наблюдалась симметрия не только на плоскости, но и в *трех* измерениях — так что в итоге получалась запрещенная икосаэдральная симметрия (Шехтман и др. [1984]). (Икосаэдральный трехмерный аналог моей плоской «плиточной» структуры с осью пятого порядка был открыт Робертом Амманном в 1975 году; см. Гарднер [1989].) Сплавы Шехтмана образовывали только крошечные микроскопические квазикристаллы, достигавшие примерно 10^{-3} мм в поперечном сечении, но позднее были найдены другие квазикристаллические вещества, в частности — алюминиево-литиево-медный сплав, у которого икосаэдрально симметричные образования могут вырастать до размеров порядка миллиметра, т. е. становятся вполне различимы невооруженным глазом (рис. 10.4).

Замечательным свойством этих квазикристаллических «плиточных» структур является то, что процесс их составления имеет существенно *нелокальный* характер. Иными словами: при построении подобного покрытия необходимо время от времени проверять состояние кристаллической решетки на расстоянии многих и многих «атомов» от места сборки, чтобы избежать серьезных ошибок при соединении составных частей. (Это чем-то напоминает то почти «сознательное

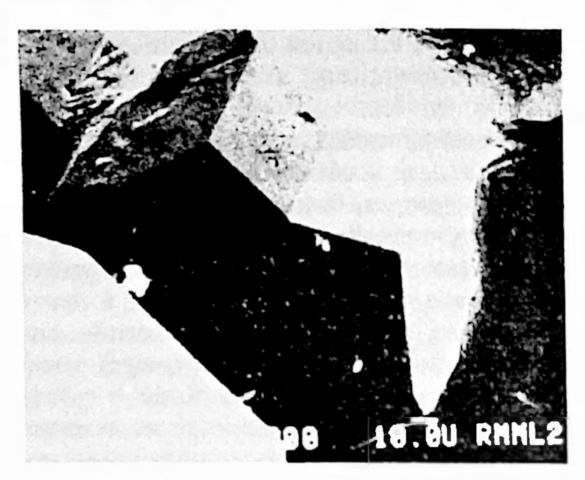


Рис. 10.4. Квазикристалл (сплав AL—Li—Cu) с, казалось бы, невозможной кристаллической симметрией. (Из Гэйл [1987].)

нащупывание», которое я связывал с естественным отбором.) Наличие такого свойства является одной из причин серьезных разногласий, возникающих сегодня в связи с вопросом о квазикристаллических структурах и их выращивании, так что было бы неразумно пытаться делать окончательные выводы до тех пор, пока не будут разрешены некоторые основополагающие проблемы. Тем не менее, никто не запрещает нам выдвигать предположения; поэтому я рискну высказать здесь свою собственную точку зрения. Во-первых, я полагаю, что некоторые из этих квазикристаллических веществ действительно имеют сложное внутреннее строение, и что расположение атомов в их структуре довольно точно повторяет строение тех плиточных структур, которыми я занимался. Во-вторых, отсюда я делаю (всего лишь гипотетическое) заключение о том, что их образование не может совершаться за счет последовательного добавления атомов, как это происходит в рамках *классической* картины роста кристаллов — но с необходимостью должна опираться на *не-локальные* и непременно квантово-механические принципы построения [7].

Механизм такого роста я представляю себе следующим образом: вместо присоединения отдельных атомов к постоянно движущейся линии роста (в случае классического роста кристаллов), происходит квантовая линейная суперпозиция большого числа различных альтернативных сочетаний

присоединяющихся атомов (путем квантовой операции U). В самом деле, согласно квантовой механике, все именно так и *должно* (почти всегда) происходить! В каждый момент времени существует не одна возможная структура, но множество альтернативных расположений атомов в сложной линейной суперпозиции. Некоторые из этих структур вырастают в гораздо более крупные образования, так что в определенный момент различия между гравитационными полями альтернативных структур превзойдут «одногравитонный предел» (или его более подходящий в данном случае аналог; см. главу 8, с. 318). На этой стадии одна из них — или, скорее, это снова будет суперпозиция, но уже в несколько урезанном виде — выделиться в качестве истинной структуры (квантовая операция R). В этот процесс роста, сопровождающийся последовательным отказом от наименее «значимых» на каждом этапе альтернатив, будут вовлекаться все бóльшее и бóльшее количество исходного вещества, пока наконец не сформируется достаточно крупный квазикристалл.

Обычно, когда природа ищет кристаллическую конфигурацию, из всех возможных она выбирает ту, которая характеризуется *наименьшим уровнем энергии* (считая фоновую температуру нулевой). Нечто аналогичное, по-моему, должно происходить и в процессе роста кристаллов, с той только разницей, что такое состояние с наименьшей энергией гораздо труднее обнаружить, а «наилучшее» расположение атомов не может быть получено просто последовательным добавлением каждый раз одного атома в надежде на то, что индивидуальному атому для этого будет достаточно решить свою *собственную* задачу минимизации. Вместо этого нам предстоит решать эту же задачу для *всей* совокупности атомов, а значит, потребуется их *совместное* усилие. Такое взаимодействие, в моем представлении, должно иметь квантово-механическую природу; и достигаться оно должно при помощи множества различных комбинаций атомных структур, которые одновременно «проверяются» в линейной суперпозиции (примерно так же, как это, вероятно, происходит в квантовом компьютере, упомянутом в конце главы 9). Условием для выбора подходящего (хотя, возможно, не лучшего) решения задачи минимизации должно быть выполнение «одногравитонного критерия» (или приемлемой в данном конкретном случае альтернативы), что, предположительно, имеет место только при соответствующих физических условиях.

Возможная связь с пластичностью мозга

Позвольте мне продолжить эти рассуждения и спросить, могут ли они иметь непосредственное отношение к процессам, происходящим в мозге. Насколько я могу судить, наиболее правдоподобно будет выглядеть связь с пластичностью мозга. Давайте вспомним, что мозг, на самом деле, похож скорее не на обычный компьютер, а на компьютер, который постоянно изменяется. За эти изменения, по-видимому, отвечают процессы активации или деактивации синапсов, которые, в свою очередь, происходят вследствие роста или сокращения дендритных шипиков (см. главу 9, с. 339; рис. 9.15). Здесь я наберусь смелости и выскажу предположение о том, что этот рост или сокращение теоретически могут подчиняться принципам наподобие тех, которые управляют квазикристаллическим ростом. При этом «тестируется» не одно из возможных альтернативных расположений, а сложная линейная суперпозиция большого числа таких расположений. До тех пор, пока эффект каждой из этих альтернатив не превышает «одногравитонного уровня» (или некоторого его аналога), они будут существовать одновременно (более того: *должны* сосуществовать, коль скоро справедливы законы U-квантовой механики). Пока не превзойден одногравитонный уровень, могут начать одновременно выполняться суперпозиции разных вычислений, что вполне соответствует принципам действия квантового компьютера. Однако, вряд ли такие суперпозиции смогут существовать достаточно долго, поскольку нервные сигналы создают электрические поля, которые должны вносить значительные возмущения в окружающую среду (хотя их миелиновые оболочки являются своего рода изоляторами). Давайте

допустим, что такие суперпозиции расчетов все-таки способны существовать в течение определенного минимального времени, которое необходимо для получения какого-нибудь *действительно* важного результата, т. е. что вплоть до этого момента «одногравитонный уровень» (или что-то подобное) в системе не достигается. Успешное завершение такого расчета будет в нашем случае той самой «целью», которая представляет собой аналог более простой «цели» минимизации энергии при квазикристаллическом росте. Таким образом, достижение этой цели будет подобно успешному росту квазикристалла!

В этих рассуждениях, конечно, много неясного и спорного, но я верю, что они описывают принципиально возможную аналогию. Рост кристалла или квазикристалла существенно зависит от концентрации нужных атомов и ионов в окрестности точек роста. Точно также можно предположить, что процессы роста или сокращения семейств дендритных шипиков, в свою очередь, находятся в прямой зависимости от степени концентрации вокруг них различных нейромедиаторов (например, таких, чья концентрация зависит от испытываемых эмоций). Какие бы расположения атомов в конце концов ни были выделены в качестве реальной структуры получившегося квазикристалла — каждый раз этому должно предшествовать решение задачи минимизации энергии. Тогда я осмелюсь по аналогии предположить, что конкретная мысль, которая возникает на поверхности мозга, тоже возникает в результате решения некоторой задачи, только на сей раз не просто задачи минимизации энергии. Эта задача будет гораздо более сложной, требующей учета желаний и намерений, которые, в свою очередь, напрямую связаны с вычислительными свойствами и функциями мозга. Я полагаю, что сознательное мышление тесно связано с отсевом тех возможных альтернатив, которые прежде входили в линейную суперпозицию. Все это имеет непосредственное отношение к неизвестным (пока!) физическим процессам, которые должны управлять пограничной областью между **U** и **R**, и которые, я уверен, будут описаны правильной теорией квантовой гравитации — ПКТГ, которую еще предстоит открыть!

Могло бы такое физическое действие быть по своей природе неалгоритмическим? Вспомним, что в общем случае задача о плиточных покрытиях, описанная в главе 4, не имеет алгоритмического решения. Можно предположить, что сходная задача в приложении к атомным структурам имеет такое же свойство «неалгоритмичности». Если эти задачи могут в принципе быть «решены» средствами, о которых я говорю, то тогда есть вероятность, что у рассматриваемого мной типа умственной деятельности действительно существует неалгоритмическая компонента. Однако для того, чтобы это было так, нам необходима определенная неалгоритмичность и в ПКТГ. Конечно, мы сейчас слишком вольно обращаемся с гипотезами — но все же приведенные выше аргументы подсказывают мне, что здесь определенно должно быть *нечто*, имеющее неалгоритмический характер.

Как быстро происходят подобные изменения в мозговых связях? На этот счет у нейрофизиологов нет единого мнения, однако, коль скоро устойчивые отпечатки в памяти могут формироваться за доли секунды, разумно предположить, что указанные изменения происходят примерно за то же время. Чтобы мои собственные идеи получили право на существование, требуется как раз примерно такая быстрота.

Временны́е задержки в реакции сознания

Теперь я хочу рассказать о двух экспериментах (описанных в работе Харта [1982]), которые проводились на добровольцах и которые, как мне кажется, имеют прямое отношение к нашим рассуждениям. В ходе этих экспериментов изучалось характерное время, которое требуется сознанию для того, чтобы осуществить определенное действие и отреагировать на внешнее воздействие. Первый из них, таким образом, изучал активную деятельность сознания, а второй — пассивную. Связанные воедино, результаты этих экспериментов представляются еще более поразительными.

Первый эксперимент был проведен Корнхубером и его коллегами в Германии,

в 1976 году (Дике и др. [1976]). Ряд испытуемых согласились на запись электрических сигналов, снятых с определенной точки поверхности головы (т. е. на электроэнцефалограмму, или ЭЭГ). Испытуемые должны были внезапно сгибать *по своему усмотрению* указательный палец. Идея заключалась в том, что на электроэнцефалограмме окажутся зафиксированы некие признаки мозговой активности, которая возникает в голове в момент принятия сознательного решения согнуть палец. Чтобы получить значимый сигнал с дорожек ЭЭГ, нужно было усреднить его по нескольким испытаниям, и результирующий сигнал при этом получался не очень «показательным». Но вот что оказалось примечательным, так это отмеченное на ленте постепенное нарастание электрического потенциала в течение *целой секунды* — а то и всех полутора — происходящее *до* того момента, как палец действительно сгибался. Это, по-видимому, означает, что сознательный процесс принятия решения занимает не менее секунды — и только затем следует его исполнение! Это очень сильно отличается по времени от куда более короткого промежутка времени реакции на внешнее раздражение, если способ реакции задан заранее (например, если испытуемый должен сгибать палец не «по собственному хотению», а исключительно в ответ на вспышку светового сигнала). В этом случае нормальной является задержка реакции длительностью около одной пятой секунды, что примерно в пять раз быстрее, чем «волевое действие», отраженное в экспериментальных данных Корнхубера (рис. 10.5).



Рис. 10.5. Эксперимент Корнхубера. Решение согнуть палец принимается за начало отсчета, однако предшествующий сигнал (усредненный по многим опытам) свидетельствует о том, что сгибание пальца «предвиделось»

Во втором эксперименте Бенджамин Либет из Калифорнийского университета, в сотрудничестве с Бертрамом Фейнстейном из Нейрологического института в Сан-Франциско (Либет и др. [1979]) проводили опыты с добровольцами, которым по независимым причинам предстояла хирургическая операция на мозге и которые соглашались на имплантацию электродов в определенные точки соматосенсорной коры своего головного мозга. Эксперимент Либета показал, что между стимуляцией кожи этих пациентов и ее осознанием проходило примерно полсекунды, несмотря на то, что сам мозг должен был получить соответствующий сигнал примерено через одну сотую долю секунды; а заранее запрограммированный «рефлексивный» ответ на этот раздражитель (см. выше) мог быть выработан мозгом примерно через одну десятую секунды (рис. 10.6). Более того: несмотря на такое «запоздалое» осознание стимулирующего воздействия, субъективное впечатление, возникавшее у пациентов, позволяло им утверждать, что не было вообще никакой задержки! (Некоторые эксперименты Либета предусматривали стимулирование зрительного бугра, см. с. 327. Результат был тот же, что и при воздействии на соматосенсорную кору головного мозга.)

Напомним, что соматосенсорная кора — это область головного мозга, куда поступают сигналы от органов чувств. Так, электрическая стимуляция зоны соматосенсорной коры, соответствующей некоторому определенному участку поверхности кожи, воспринимается испытуемым в точности так, как если бы что-то действительно коснулось его кожи в этом месте. Однако оказалось, что если эта электрическая стимуляция длится менее, чем примерно полсекунды, то испытуемый не получает никакого ощущения вообще. В противоположность этому, прямая стимуляция точки на самой поверхности кожи может ощущаться даже в том случае, когда она мгновенна.

Теперь предположим, что сначала тронули кожу, а затем подвергли электрической стимуляции соответствующую точку в соматосенсорной коре. Что чувствует пациент? Если электростимуляция производится примерно через четверть секунды после касания

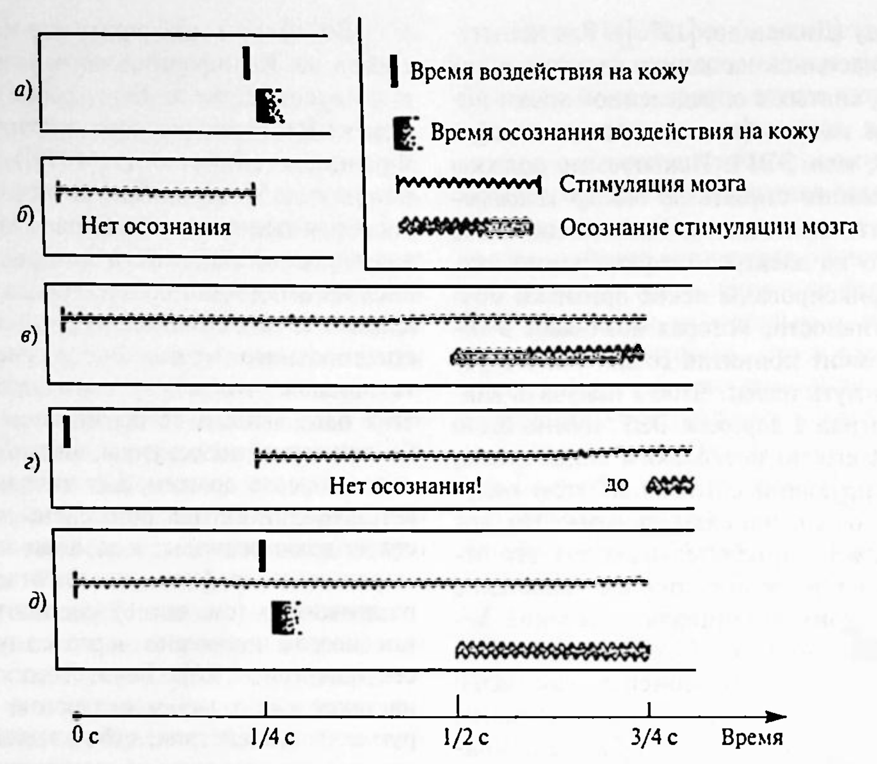


Рис. 10.6. Эксперимент Либета: *а*) воздействие на кожу, «как кажется», осознается практически без задержки; *б*) стимуляция коры, длящаяся меньше полсекунды, не осознается; *в*) стимуляция коры продолжительностью более, чем полсекунды, осознается примерно через полсекунды; *г*) такая стимуляция коры может приводить к «обратной маскировке» более раннего воздействия на кожу, показывая, что осознание этого воздействия, на самом деле, еще *не имело места* к моменту стимуляции коры; *д*) если стимуляция кожи произведена вскоре *после* такого воздействия на кору, то тогда реакция на стимуляцию кожи «сдвигается назад», хотя с реакцией на стимуляцию соматосенсорной коры этого не происходит

кожи, тогда касание кожи не ощущается вообще! Этот эффект получил название *обратной маскировки*. Каким-то образом стимулирование коры мозга препятствует осознанию обычного ощущения от касания кожи. Сознательное восприятие может быть предотвращено («замаскировано») более поздним событием, если это событие происходит не позднее, чем через полсекунды. Это говорит о том, что сознательное восприятие такого воздействия проявляется примерно через полсекунды после самого воздействия!

Однако, по все видимости, испытуемые не осознавали столь долгой задержки своего восприятия. Найти разумное объяснение этому удивительному открытию можно было бы, предположив, например, что «время» всех наших «восприятий» действительно отличается примерно на полсекунды от «реального времени» — как будто наши внутренние часы просто «неверны» и отстают на полсекунды или около того. Время восприятия некоторого события в этом случае должно всегда *отстоять* на полсекунды от того момента, когда указанное событие произошло. Это дало бы возможность представить последовательную, хотя и неудовлетворительно замедленную картину нашего чувственного восприятия.

Возможно, что-то подобное происходило во второй части эксперимента Либета, где он *вначале* стимулировал кору мозга, продолжая эту стимуляцию гораздо дольше, чем полсекунды — и касался кожи во время этой стимуляции, но ранее, чем через полсекунды после ее начала. И стимуляция коры, и касание кожи воспринимались раздельно, так что пациенту было ясно, какой именно стимул он воспринимал. Однако, когда его спрашивали, какой стимул он ощутил

первым, он обычно отвечал, что это было касание кожи, тогда как, на самом деле, стимуляция коры всегда предшествовало ему! Таким образом, испытуемый и вправду сдвигал свое восприятие прикосновения к коже *назад во времени* примерно на полсекунды (см. рис. 10.6). Однако это представляется не просто глобальной «ошибкой» внутреннего ощущения времени, но более тонкой структурной перестройкой временной шкалы восприятия событий. Поскольку в случае стимуляции коры (учитывая, что она в действительности воспринимается не позднее, чем через полсекунды после ее начала), такая задержка *не* наблюдается.

Опираясь на данные первого из вышеописанных экспериментов, мы, по всей видимости, можем сделать вывод о том, что сознательному действию необходимо примерно около секунды или полутора секунд на то, чтобы быть приведенным в исполнение; а в соответствии со вторым экспериментом — что осознание внешнего события, по-видимому, не происходит раньше, чем через полсекунды после момента события. Представим себе, что происходит, когда человек реагирует на некоторое неожиданное внешнее событие. Предположим также, что ответ требует моментального сознательного действия. Если принять в расчет открытие Либета, то должно пройти полсекунды прежде, чем сознание «включится»; и после этого, как следует из опытов Корнхубера, потребуется еще секунда, а то и более, прежде чем человек «осознанно» отреагирует на это событие. Таким образом, весь процесс — от сенсорного восприятия до моторного отклика — занимает примерно две секунды! Очевидный вывод из этих двух экспериментов, если рассматривать их вместе, напрашивается сам собой: сознание *вообще* не может быть задействовано там, где ответная реакция на внешнее событие должна занимать не более пары секунд!

Странная роль времени в сознательном восприятии

Можно ли доверять результатам этих экспериментов? Если это так, мы с необходимостью приходим к выводу, что мы действуем как «автоматы», когда чтобы изменить реакцию, требуется менее одной или двух секунд. Становится несомненным, что сознание, по сравнению с другими механизмами нервной системы, работает довольно медленно. Я и сам замечал, как иной раз моя рука продолжала захлопывать дверцу машины еще несколько мгновений спустя после того, как я заметил, что в машине осталось что-то нужное, и как сознательный приказ остановить движение руки патологически не успевает за ним — так медленно все это происходит. Но требует ли это и впрямь целую секунду, а то и две? Такой длительный промежуток времени кажется мне невероятным. Разумеется, мое *сознательное* восприятие забытой вещи в машине вместе с моим воображаемым «свободно-волевым» приказом остановить руку вполне *могли* случиться уже после обоих этих событий. Возможно, сознание — это просто наблюдатель, который воспринимает происходящее не иначе, как повторное исполнения всего спектакля. Аналогично, опираясь на результаты вышеописанных опытов, можно считать, что какая бы роль не отводилась сознанию, например, при отбивании теннисного мяча — не говоря уже об игре в пинг-понг — у него просто не хватило бы времени на ее исполнение! Несомненно, что у опытных игроков все наиболее важные приемы игры наиподробнейшим образом запрограммированы на уровне мозжечка. Но вот чтобы сознание не играло *никакой* роли в принятии решения относительно того, какой удар должен быть выполнен в конкретный момент — в это я могу поверить с большим трудом. Конечно, многое заложено в интуитивном угадывании следующего движения соперника, и наличии множества заранее просчитанных и приготовленных вариантов ответа на каждое из них — но такой сценарий, при котором сознание *вообще* не участвует в формировании ответной реакции, кажется мне неэффективным и маловероятным. Эти возражения были бы еще более уместны в случае обычного разговора. Здесь также собеседники могут частично догадываться, что скажет другой, но в ответах оппонента должно достаточно часто присутствовать что-то неожиданное, иначе беседа просто потеряла бы смысл! И вряд ли кто будет

спорить, что в обычном разговоре на то, чтобы ответить собеседнику, требуется куда меньше, чем две секунды времени!

Похоже, есть основания сомневаться, что в экспериментах Корнхубера сознанию «действительно» нужно полторы секунды для выполнения задуманного действия. Хотя *усредненная* по всем записям ЭЭГ задержка между возникновением намерения согнуть палец и непосредственным действием дает как раз такую величину, тем не менее может оказаться, что только в *некоторых* случаях намерение проявлялось столь рано — причем часто *не* приводило в действительности к сгибанию пальца; тогда как во многих других случаях сознание приводило палец в движение гораздо быстрее. (На самом деле, более поздние эксперименты, см. Либет [1987, 1989], позволяют сделать иные, чем у Корнхубера, выводы. Однако загадки, связанные с временны́ми аспектами сознания, так и остались нерешенными.)

Давайте представим себе на минуту, что результаты обоих экспериментов *справедливы*. Тогда с необходимостью следует признаться в том, что мы могли до сих пор идти по глубоко ошибочному пути, используя при изучении работы сознания обычные физические правила для *времени*! В самом деле, есть нечто весьма странное в том, как время входит в наше сознательное восприятие, и я думаю, что для интерпретации этого феномена в рамках наших традиционных представлений может понадобиться совсем иная концепция. Сознание — это, в конце концов, единственное известное нам явление, согласно которому время «течет»! Способ рассмотрения времени в современной физике не отличается по существу от способа рассмотрения *пространства*[6]; так что, на самом деле, «время» в физических процессах не «течет» — вместо этого рассматривается статичное «пространство-время», где фиксируются события, происходящие в нашей вселенной! Однако, мы воспринимаем время *текущим* (см. главу 7). Я полагаю, что и здесь присутствует некая иллюзия, и что, *на самом деле*, время нашего восприятия не течет линейно в одном направлении (что бы это ни значило!). «Кажущуюся» временную упорядоченность воспринимаемых событий мы, по-моему, привносим в наши ощущения сами для того, чтобы как-то согласовать их с единым для окружающего нас физического мира поступательным движением во времени.

Некоторые могли бы усмотреть в подобных замечаниях изрядную долю беспочвенного «философствования» — и их обвинения, конечно же, были бы справедливы. Как можно «ошибаться» относительно того, что ты действительно воспринимаешь? Ясно, что ощущения — это (*по определению*) то, что непосредственно осознается; поэтому они просто не могут быть «неправильными». Тем не менее, я думаю, что, на самом деле, мы все-таки «*ошибаемся*», когда воспринимаем время, как движущееся вперед — несмотря на неадекватность доступных мне языковых средств для описание моего убеждения; и что существуют свидетельства, подтверждающие справедливость такой гипотезы (см. Черчланд [1984]).

Исключительно ярким примером (см. с. 362) является способность Моцарта «охватывать единым взглядом» всю музыкальную композицию, даже когда «она бывает довольно длинной». Исходя из описания самого Моцарта, можно предположить, что этот «взгляд» охватывал все существенные стороны произведения — и что, тем не менее, интервал времени (в обычном физическом смысле), необходимый для подобного сознательного восприятия композиции, оказывался заведомо короче того, который потребовался бы для ее исполнения. Кто-то может считать, что все воспринималось совсем по-другому, и Моцарт «видел» свое будущее произведение в форме пространственно-распределенных образов или, допустим, готовой музыкальной партитуры. Но и для внимательного прочтения партитуры таких размеров необходимо довольно много времени — и, к тому же, я сильно сомневаюсь

---

[6] Эта симметрия между временем и пространством становится еще более удивительной в случае *двумерного* пространства-времени. Уравнения двумерной физики пространства-времени оказываются существенно симметричны относительно взаимозамены координат пространства и времени — однако, в двумерной физике никто не стал бы требовать от пространства, чтобы оно «текло». Трудно поверить, что «реальное течение» времени в нашем восприятии окружающего мира обусловлено разве что асимметрией между числом измерений пространства (3) и измерений времени (1), характерной для нашего пространства-времени.

в том, что исходное восприятие Моцартом своей композиции могло принимать указанную форму (иначе он бы наверняка об этом сказал!). Образное восприятие кажется более вероятным; однако, (как и в большинстве случаев визуализации в математике, с которыми я лично сталкивался) я сильно сомневаюсь, что в сознании Моцарта мог совершаться прямой перевод музыки на язык зрительных образов. Мне кажется, что интерпретировать «взгляд» Моцарта правильнее всего с чисто *музыкальной* точки зрения, с четким временны́м распределением, которое обычно возникает при прослушивании (или исполнении) музыкального произведения. Ведь музыка состоит из звуков, воспроизведение которых требует определенного времени — *времени*, которое, со слов самого Моцарта, «...позволяет мое воображение».

Послушайте четырехчастную фугу из последнего раздела *Искусства фуги* И. С. Баха. Знатоки Баха не могут не переживать стресс, когда музыка останавливается после десяти минут звучания, сразу же после вступления третьей темы. Кажется, что композиция каким-то образом все еще существует «там, вовне» — просто сейчас она внезапно замерла. Бах покинул этот мир, не успев закончить свою работу и не оставив нам ни единого намека на то, как он намеревался продолжить ее. Однако, она начинается с такой уверенностью и бесспорным мастерством, что невозможно представить себе, чтобы у Баха в то время не было ясного представления о всех ключевых моментах своего будущего произведения. Нужно ли ему было мысленно исполнять композицию в обычном темпе, каждый раз «проигрывая» ее заново по мере возникновения новых идей и различных поправок? Я не могу себе представить, что это происходило таким образом. Как и Моцарт, он должен был представлять себе работу целиком, связывая в голове воедино как ее сложнейшую структуру, так и многочисленные замысловатые украшения — все то, без чего не мыслимо создание фуг. При том, что и временны́е характеристики музыки важны никак не меньше. Ибо как музыка может оставаться музыкой, если она не исполняется «в реальном времени»?

Рождение замысла романа или истории можно было бы рассматривать как аналогичный (хотя, на первый взгляд, и менее непостижимый) процесс. Охватывая внутренним взором всю жизнь персонажа, необходимо продумывать различные события, которые автор, как кажется, просто не сможет вставить в сюжет, не проиграв предварительно в «реальном времени». Однако это далеко не всегда необходимо. Даже сохранившиеся в памяти впечатления от лично пережитых событий оказываются настолько «сжатыми», что их можно мысленно «пережить» вновь за доли секунды!

Видимо, существует определенное (и при том значительное) сходство между сочинением музыки и математическим мышлением. Многие, вероятно, уверены, что математическое доказательство строится в виде цепочки последовательных утверждений, где каждый шаг вытекает из предыдущего. Но, на самом деле, замысел доказательства едва ли когда возникает подобным образом. Общее представление и лишь интуитивно понятное концептуальное содержание — вот что в действительности необходимо для построения математического доказательства; и это едва ли можно соотнести с тем временем, которое потребовалось бы в дальнейшем для его полного последовательного изложения.

Предположим далее, что мы допускаем отсутствие соответствия между внутренней шкалой времени нашего сознания — с одной стороны, и течением времени в окружающем нас физическом мире — с другой. Не рискуем ли мы при этом столкнуться с парадоксом? Предположим к тому же, что в природе наших сознательных действий заложено что-то неуловимо телеологическое, позволяющее будущим впечатлениям от действия в прошлом оказывать влияние на само это действие. Ясно, что *это* могло бы привести нас к противоречию, подобному парадоксальным следствиям из предположения о возможности распространения сигнала со скоростью, превышающей скорость света, которое мы рассматривали — и совершенно обоснованно отвергли — в конце главы 5 (см. с. 195). Я считаю, что никакого парадокса здесь быть *не* должно — как это непосредственно

следует из моих утверждений, касающихся самого понятия сознания и его возможностей. Если вы помните, я выдвигал предположение о том, что сознание, в сущности, есть способность «вúдеть» непреложную истину; и что оно может представлять собой своеобразный контакт с миром идеальных математических идей Платона. Напомню, что мир Платона сам по себе имеет вневременную природу. Восприятие истины Платона не несет подлинной информации — имея в виду технический аспект понятия «информации», связанный с возможностью ее передачи; так что, на самом деле, не будет никакого противоречия даже в том случае, если бы подобное сознательное восприятие распространялось обратно во времени!

Но даже если мы согласимся с тем, что сознание связанно со временем таким причудливым образом — и что благодаря сознанию происходит своего рода контакт между нашим физическим миром и определенной вневременной сущностью — как тогда быть с физически обусловленным и упорядоченным во времени действием материального мозга? И снова мы, по-видимому, вынуждены отводить сознанию роль простого «зрителя» — в противном случае нам придется так или иначе подтасовывать физические законы, чтобы не нарушить естественное развитие событий. Однако я все же *отстаиваю* активную роль сознания, которая дает ему преимущество в ходе естественного отбора. Ответ на эту дилемму, как мне кажется, может заключаться в том странном способе, как должна действовать ПКТГ, разрешая конфликт между двумя квантово-механическими процедурами **U** и **R** (см. с. 306, 317).

Вспомним о проблемах со временем, которые возникали в результате наших попыток согласовать **R**-процедуру со (специальной) теорией относительности (главы 6, 8, с. 252, 321). При описании этой процедуры в обычных пространственно-временных терминах она, кажется, вообще теряет всякий смысл. Рассмотрим квантовое состояние пары частиц. Как правило, такое состояние должно быть *коррелированным* (т. е. описываться *не* простым выражением $|\psi\rangle|\chi\rangle$, где каждый из сомножителей $|\psi\rangle$ и $|\chi\rangle$ описывает только одну частицу, но представляет собой сумму вида $|\psi|\chi\rangle + |\alpha|\beta\rangle + \ldots + |\varrho|\sigma\rangle$). Тогда наблюдение за одной из частиц окажет на другую нелокальное воздействие, которое не может быть описано в обычных пространственно-временных терминах, согласующихся со специальной теорией относительности (ЭПР; эффект Эйнштейна—Подольского—Розена). Подобные нелокальные эффекты должны неявным образом присутствовать в предложенной мною «квазикристаллической» аналогии для роста и сокращения дендритов.

Под «наблюдением» я здесь понимаю усиление действия каждой наблюдаемой частицы до тех пор, пока не достигается некий уровень, соответствующий, например, «одногравитонному критерию» в рамках ПКТГ. В более «стандартной» терминологии «наблюдение» — это крайне нечетко определенное понятие; и, согласитесь, трудно себе представить, как можно начинать теоретические исследования в области квантово-механического описания работы мозга, если приходится считаться с необходимостью рассматривать мозг в качестве объекта, который постоянно «наблюдает сам себя»!

Моя собственная идея заключается в том, что концепция ПКТГ, напротив, должна дать нам *объективную* физическую теорию редукции вектора состояния (**R**-процедуры), которая никак *не* будет зависеть от нашего взгляда на сознание. У нас пока нет такой теории, но, по крайней мере, можно быть уверенным, что ее создатели не будут спотыкаться о фундаментальные вопросы, связанные с точным определением сознания!

Я полагаю, что именно *после* открытия теории ПКТГ у нас в конце концов появится возможность описания с ее помощью феномена сознания. Вообще говоря, я склонен считать, что априори предполагаемые свойства ПКТГ окажутся на самом деле еще менее удобными для их адекватного описания в обычных пространственно-временных терминах, чем упомянутые выше загадочные явления ЭПР в системе двух частиц. Если я прав, и сознание напрямую связано с будущей теорией ПКТГ, то лишь с большой натяжкой нам удастся применить к нему наши привычные пространственно-временные описания.

Заключение: точка зрения ребенка

В этой книге я привел множество доводов, призванных показать несостоятельность точки зрения — как выясняется, одной из наиболее распространенных в современной философии — согласно которой наше мышление в основе своей идентично действию очень сложного компьютера. Когда в явном виде делается предположение о том, что простое выполнение алгоритма может привести к возникновению *осознанного восприятия*, то используется терминология концепции «сильного ИИ» Серла. В другие термины (такие, как «функционализм») подчас вкладывают более широкий смысл.

Некоторые читатели, возможно, с самого начала считали «сторонников сильного ИИ» законченной деревенщиной! Разве не «очевидно», что обычные вычисления не могут вызвать удовольствие или причинить боль; что они не способны понимать поэзию, наслаждаться красотой вечернего неба или магией звуков; что они не могут надеяться, любить или отчаиваться; что у них не может возникнуть настоящей независимой цели существования? Однако наука, кажется, заставляет нас поверить в то, что все мы — просто ничтожные частички мира, полностью управляемого (пусть даже только вероятностно) очень точными математическими законами. Равно как и наш мозг, который, казалось бы, контролирует все наши действия, но, в свою очередь, точно также подчиняются тем же самым законам. Возникает ощущение, что вся эта скоординированная физическая активность является, на самом деле, ничем иным, как выполнением некоторого всеобъемлющего (возможно, по своей природе вероятностного) вычислительного процесса — и, следовательно, наш мозг и наш разум нужно рассматривать исключительно в терминах такого вычисления. Может быть, когда степень сложности подобного алгоритма становится чрезвычайно высокой, он приобретает те поэтические или субъективные качества, которые мы привыкли ассоциировать с понятием «разум». Однако трудно избавиться от навязчивого ощущения, что в такой картине всегда будет чего-то не хватать.

В своих рассуждениях я пытался найти обоснование своей уверенности в том, что и вправду должно быть нечто важное и существенное, остающееся за рамками любой «алгоритмической» картины мира. Тем не менее, я по-прежнему связываю свои надежды на разгадку тайны разума с наукой в целом, и математикой в частности. Здесь возникает очевидная дилемма — однако, я старался показать, что из этой ситуации *есть* совершенно естественный выход. Свойство вычислимости — не то же самое, что математическая точность. Сколько тайны и красоты в точном математическом мире Платона — а ведь бóльшая непознанная часть этого мира связана с понятиями, которые находятся за пределами той сравнительно небольшой его части, где располагаются алгоритмы и вычисления.

Сознание представляется мне таким важным явлением, что я просто не могу поверить в возможность его «случайного» возникновения в результате сколь угодно сложных вычислений. Ведь именно благодаря ему мы можем говорить о самом существовании вселенной. Некоторые считают, что вселенная, законы которой не допускают зарождение сознания, вообще не является вселенной. Я бы даже сказал, что все математические описания вселенной, которые до сих пор были сделаны, не должны удовлетворять этому критерию. Только сознание могло вызвать предполагаемую «теоретическую» вселенную к жизни!

Некоторые доводы, приведенные мной в этих главах, могут показаться чересчур сложными для понимания, другие представляются слишком спорными — хотя, по-моему, немало здесь и таких, которые, наоборот, никак нельзя оставить без внимания. Тем не менее, за всеми этими техническими рассуждениями стоит одно — ощущение «очевидности» предположения о том, что разум, наделенный *сознанием*, просто не может работать подобно компьютеру, несмотря на алгоритмическую природу многих составляющих нашей умственной деятельности.

Это тот тип очевидности, который доступен и ребенку — хотя со временем этот

ребенок, став уже взрослым, будет вынужден поверить в то, что очевидные проблемы — это «не проблемы», что они могут быть сведены на нет при помощи тщательно подобранных рассуждений и удачных определений. Дети иногда ясно видят многие вещи, которые в более зрелые годы теряют для них свою очевидность. Мы часто забываем чувство восхищения, которое мы испытывали в детстве, когда впоследствии на наши плечи ложится груз повседневных забот «мира взрослых». Дети не боятся задавать самые элементарные вопросы из числа тех, которые нам, взрослым, задавать уже «стыдно». Что происходит с каждым из потоков нашего сознания после смерти; где было сознание до нашего рождения; могли бы мы стать, или уже были, кем-то еще; почему мы вообще воспринимаем мир; почему мы здесь; и почему, в конце концов, есть такая вселенная, в которой мы можем существовать? Это загадки, которые имеют обыкновение возникать в момент пробуждения способности осознавать в каждом из нас — и, несомненно, с первыми проблесками подлинного самосознания в любом живом существе.

Я помню свои собственные попытки разрешить для себя многие из этих загадок, когда я был ребенком. Допустим, думал я, мое сознание имеет возможность внезапно поменяться с чьим-то другим — как в таком случае я могу быть уверен, что нечто подобное уже не произошло со мной раньше, предполагая, что каждый человек хранит в памяти только то, что относится к нему лично? Как я мог бы тогда объяснить такой опыт «обмена» кому-то еще? Или все это вообще не имеет никакого смысла? А что если я просто проживаю те же самые десять минут жизни снова и снова — и каждый раз с одними и теми же впечатлениями? Может быть, для меня «существует» только настоящее? Может быть, «я» завтрашнего или вчерашнего дня — это в действительности совершенно иная личность с независимым сознанием? Может быть, я на самом деле живу «задом наперед» во времени, и мой поток сознания направлен в прошлое, так что моя память говорит не о том, что уже *произошло* со мной, но о том, что еще только *должно произойти* — и неприятности в школе еще впереди и, к сожалению, уже не за горами? Есть ли вообще какое-нибудь значимое различие между *таким* и обычным течением времени, которое позволило бы считать одно из них «правильным», а второе — нет? Для того, чтобы иметь принципиальную возможность получать ответы на подобные вопросы, необходима теория сознания. Но как можно даже *начинать* объяснять сущность таких проблем тому, кто сам не обладает сознанием?..

Примечания

1. Как мы видели в главе 4 (с. 125), проверка справедливости доказательства в формальной системе всегда имеет алгоритмический характер. И наоборот, любой алгоритм, который позволяет получать математически истинные утверждения, всегда можно добавить в систему аксиом и правил вывода обычной логики («предикатного исчисления»), тем самым создавая новую формальную систему выведения математических истин.
2. Некоторых читателей может беспокоить тот факт, что в среде математиков действительно существуют различные точки зрения. Вспомним рассуждения, приведенные в главе 4. Однако имеющиеся разногласия не так важны для нас. Они относятся только к в высшей степени абстрактным вопросам, касающимся очень больших множеств, в то время как мы вполне можем ограничиться утверждениями арифметического характера (с конечным числом кванторов существования и всеобщности) и применить дальнейшие рассуждения. (Возможно, здесь допущено некоторое преувеличение, поскольку принцип рефлексии, относящийся к бесконечным множествам, может иногда использоваться для вывода утверждений в арифметике.) Что касается крайне догматичного и не желающего соглашаться с Геделем формалиста, для которого такая вещь, как математическая истина, вообще *не существует*, то я его буду просто-напросто игнорировать, поскольку он явно не обладает

способностью интуитивного понимания истины, которой посвящены наши рассуждения!

Конечно, математики иногда допускают ошибки. Кажется, сам Тьюринг считал, что именно *это* и есть «лазейка», которая позволяет обойти аргументы геделевского типа в пользу того, что человеческое мышление существенно неалгоритмично. Но лично мне кажется невероятным, что свойство людей ошибаться каким-либо образом связано с нашей способностью к прозрениям! (Между прочим, *генераторы* случайных чисел могут быть успешно реализованы при помощи алгоритмов.)

3. Термин «черная дыра» вошел во всеобщее употребление много позже, около 1968 года (главным образом благодаря пророческим идеям американского физика Джона А. Уилера).

4. Мне кажется, что потребность животных во сне, во время которого они иногда *видят сны* (как это бывает часто заметно у собак), может служить свидетельством того, что они, вполне вероятно, наделены сознанием. Ибо разница между сном без сновидений и сном со сновидениями, по-видимому, во многом определяется как раз наличием сознания.

5. В случае специальной или общей теории относительности под «временами» следует понимать «одновременные пространства» или «пространственно-подобные поверхности» (с. 187, 197).

6. Однако в случае пространственно-бесконечной вселенной есть затруднения, поскольку тогда возникает (как и в случае множественных миров) бесконечное количество копий наблюдателя и его непосредственного окружения! Будущее поведение каждой копии может несколько отличаться, и никто не в состоянии сказать наверняка, какой из приблизительных копий самого себя, смоделированных математическим путем, он мог бы на самом деле «быть»!

7. Даже в ходе реального роста некоторых кристаллов могут возникать подобные проблемы — например, там, где исходная клетка кристаллической решетки содержит несколько сот атомов (случай так называемых «фаз Фрэнка—Каспера»). С другой стороны, следует упомянуть, что теоретический «почти локальный» (хотя все же нелокальный) процесс роста квазикристаллов с осью пятого порядка был предложен Онодой, Стайнхардтом, Ди Винченцо и Соколаром [1988].

Эпилог

«...СЕБЯ ЧУВСТВУЕШЬ? О... весьма интересный вопрос, мой мальчик... э-э... я и сам хотел бы знать ответ», — сказал Главный конструктор. — «Давайте посмотрим, что может сказать наш друг об этом... странно... э-э... *Ультроник* говорит, что он не понимает, что... он не может даже понять, что ты имеешь в виду!» Отдельные смешки в аудитории переросли в громовой хохот. Адам чувствовал себя крайне неловко. Они могли отреагировать как угодно, но только не смеяться.

Литература

Ааронов, Альберт [1981] Aharonov, Y. and Albert, D. Z. (1981). Can we make sense out of the measurement process in relativisitic quantum mechanics? *Phys. Rev.*, D24, 359–70.

Ааронов, Бергманн, Лебовиц [1964] Aharonov, Y., Bergmann, P., and Lebowitz, J. L. (1964). Time symmetry in the quantum process of measurement. Опубликовано в *Quantum theory and measurement* (ed. J. A. Wheeler and W. H. Zurek), Princeton University Press, 1983; первоначально опубликовано в *Phys. Rev.*, 134B, 1410–16.

Адамар [1945] Hadamard, J. (1945). *The psychology of invention in the mathematical field.* Princeton University Press. (Рус. пер.: *Адамар Ж.* Исследования психологии процессов изобретательства в области математики. М.: Педагогика, 1970.)

Аспект, Гранжьер [1986] Aspect, A. and Grangier, P. (1986). Experiments on Einstein—Podolsky—Rosen-type correlations with pairs of visible photons. Опубликовано в *Quantum concepts in space and time* (ed. R. Penrose and C. J. Isham), Oxford University Press.

Аштекар, Балашандран, Санг [1989] Ashtekar, A., Balachandran, A. P., and Sang Jo (1989). The CP problem in quantum gravity. *Int. J. Mod. Phys.*, A6, 1493–514.

Барба [1989] Barbour, J. B. (1989). *Absolute or relative motion?* Volume 1: *The discovery of dynamics.* Cambridge University Press, Cambridge.

Барроу [1988] Barrow, J. D. (1988). *The world within the world.* Oxford University Press.

Барроу, Типлер [1986] Barrow, J. D. and Tipler, F. J. (1986). *The anthropic cosmological principle.* Oxford University Press.

Бекенштейн [1972] Bekenstein, J. (1972). Black holes and entropy. *Phys. Rev.*, D7, 2333–46.

Белинский, Халатников, Лифшиц [1970] Belinskii, V. A., Khalatnikov, I. M., and Lifshitz, E. M. (1970). Oscillatory approach to a singular point in the relativistic cosmology. *Adv. Phys.* 19, 525–73.

Белинфанте [1975] Belinfante, F. J. (1975). *Measurement and time reversal in objective quantum theory.* Pergamon Press, New York.

Белл [1987] Bell, J. S. (1987). *Speakable and unspeakable in quantum mechanics.* Cambridge University Press.

Бенасерраф [1967] Benacerraf, P. (1967). God, the Devil and Gödel. *The Monist*, 51, 9–32.

Блэйкмор, Гринфилд [1987] Blakemore, C. and Greenfield, S. (eds.) (1987). *Mindwaves: thoughts on intelligence, identity and consciousness.* Basil Blackwell, Oxford.

Блюм, Шуб, Смэйл [1989] Blum, L., Shub, M., and Smale, S. (1989). On a theory of computation and complexity over the real numbers: NP completeness, recursive functions and universal machines. *Bull. Amer. Math. Soc.*, 21, 1–46.

Бови [1982] Bowie, G. L. (1982). Lucas' number is finally up. *J. of Philosophical Logic*, 11, 279–85.

Бом [1951] Bohm, D. (1951). The paradox of Einstein, Rosen and Podolsky. Опубликовано в *Quantum theory and measurement* (ed. J. A. Wheeler and W. H. Zurek), Princeton University Press, 1983; первоначально опубликовано в *Quantum theory*, D. Bohm, Ch. 22, sect. 15–19. Prentice-Hall, Englewood-Cliffs.

Бом [1952] Bohm, D. (1952). A suggested interpretation of the quantum theory in terms of 'hidden' variables, I and II. Опубликовано в *Quantum theory and measurement* (ed. J. A. Wheeler and W. H. Zurek), Princeton University Press, 1983; первоначально опубликовано в *Phys. Rev.*, 85, 166–93.

Бонди [1960] Bondi, H. (1960). Gravitational waves in general relativity. *Nature* (*London*), 186, 535.

Брукс, Мательски [1981] Brooks, R. and Matelski, J. P. (1981). The dynamics of 2-generator subgroups of PSL(2,C), Riemann surfaces and related topics: *Proceedings of the* 1978 *Stony Brook Conference*, edited by I. Kra and B. Maskit, *Ann. Math Studies*, 97. Princeton University Press, Princeton.

Бэйлор, Лам, Яу [1979] Baylor, D. A., Lamb, T. D., and Yau, K.-W. (1979). Responses of retinal rods to single photons. *J. Physiol.*, 288, 613–34.

Вайнберг [1977] Weinberg, S. (1977). *The first three minutes: A modern view of the origin of the universe.* André Deutsch, London. (Рус. пер.: *Вайнберг С.* Первые три минуты. М.: Энергоиздат, 1981.)

Вайскранц [1987] Weiskrantz, L. (1987). Neuropsychology and the nature of consciousness. Опубликовано в *Mindwaves* (ed. C. Blakemore and S. Greenfield), Blackwell, Oxford.

Валтц [1982] Waltz, D. L. (1982). Artificial intelligence. *Scientific American*, 247 (4), 101–22.

Вестфолл [1980] Westfall, R. S. (1980). *Never at rest.* Cambridge University Press.

Вигнер [1960] Wigner, E. P. (1960). The unreasonable effectiveness of mathematics. *Commun. Pure Appl. Math.*, **13**, 1–14. (Рус пер.: *Вигнер Э.* Непостижимая эффективность математики в естественных науках // Вигнер Э. Инвариантность и законы сохранения. Этюды о симметрии. М.: УРСС, 2002. С. 182–198.)

Вигнер [1961] Wigner, E. P. (1961). Remarks on the mind-body question. Опубликовано в *The scientist speculates* (ed. I. J. Good), Heinemann, London. Перепечатано в E. Wigner (1967), *Symmetries and reflections*, Indiana University Press, Bloomington, и в *Quantum theory and measurement* (ed. J. A. Wheeler and W. H. Zurek), Princeton University Press, 1983.

Вильсон, Ривз, Газзанига, Калвэр [1977] Wilson, D. H., Reeves, A. G., Gazzaniga, M. S., and Culver, C. (1977). Cerebral commissurotomy for the control of intractable seizures. *Neurology*, 27, 708–15.

Виноград [1972] Winograd, T. (1972). Understanding natural language. *Cognitive Psychology*, **3**, 1–191. (Рус. пер.: *Виноград Т.* Программа, понимающая естественный язык. М.: Мир, 1976.)

Вуттерс, Цурек [1982] Wootters, W. K. and Zurek, W. H. (1982). A single quantum cannot be cloned. *Nature*, **299**, 802–3.

Газзанига [1970] Gazzaniga, M. S. (1970). *The bisected brain*. Appleton-Century-Crofts, New York.

Газзанига, Леду, Вильсон [1977] Gazzaniga, M. S., LeDoux, J. E., and Wilson, D. H. (1977). Language, praxis, and the right hemisphere: clues to some mechanisms of consciousness. *Neurology*, **27**, 1144–7.

Галилей [1638] Galilei, G. (1638). *Dialogues concerning two new sciences*. Macmillan edn 1914; Dover Inc. (Рус. пер.: *Галилей Г.* Диалог о двух главнейших системах мира — Птолемеевской и Коперниковой. М.—Л.: Гостехиздат, 1948.)

Ганди [1988] Gandy, R. (1988). The confluence of ideas in 1936. Опубликовано в *The universal Turing machine: a half-century survey* (ed. R. Herken), Kammerer & Unverzagt, Hamburg.

Гарднер [1958] Gardner, M. (1958). *Logic machines and diagrams*. University of Chicago Press.

Гарднер [1983] Gardner, M. (1983). *The whys of a philosophical scrivener*. William Morrow and Co., Inc., New York.

Гарднер [1989] Gardner, M. (1989). *Penrose tiles to trapdoor ciphers*. W. H. Freeman and Company, New York. (Рус. пер.: *Гарднер М.* От мозаик Пенроуза к надежным шрифтам. М.: Мир, 1993.)

Гедель [1931] Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, **38**, 173–98.

Герох, Хартли [1986] Geroch, R. and Hartle, J. B. (1986). Computability and physical theories. *Found. Phys.*, **16**, 533.

Гирарди, Римини, Вебер [1980] Ghirardi, G. C., Rimini, A., and Weber, T. (1980). A general argument against superluminal transmission through the quantum mechanical measurement process. *Lett. Nuovo. Chim.*, **27**, 293–8.

Гирарди, Римини, Вебер [1986] Ghirardi, G. C., Rimini, A., and Weber, T. (1986). Unified dynamics for microscopic and macroscopic systems. *Phys. Rev.*, **D34**, 470.

Грэгори [1981] Gregory, R. L. (1981). *Mind in science; A history of explanations in psychology and physics*. Weidenfeld and Nicholson Ltd.

Грюнбаум, Шепард [1981] Grünbaum, B. and Shephard, G. C. (1981). Some problems on plane tilings. Опубликовано в *The mathematical Gardner* (ed. D. A. Klarner), Prindle, Weber and Schmidt, Boston. (Рус. пер.: *Грюнбаум Б., Шепард Дж. Ч.* Некоторые проблемы, связанные с плоскими мозаиками // Математический цветник: Сборник статей и задач. М.: Мир, 1983.)

Грюнбаум, Шепард [1987] Grünbaum, B. and Shephard, G. C. (1987). *Tilings and patterns*. W. H. Freeman.

Гуд [1969] Good, I. J. (1969). Gödel's theorem is a red herring. *Brit. J. Philos. Sci.*, **18**, 359–73.

Гэйл [1987] Gayle, F. W. (1987). Free-surface solidification habit and point group symmetry of a faceted icosahedral Al-Li-Cu phase. *J. Mater. Res.*, **2**, 1–4.

де Бройль [1956] de Broglie, L. (1956). *Tentative d'interprétation causale et nonlinéaire de la mécanique ondulatoire*. Gauthier-Villars, Paris.

Де Витт, Грэхем [1973] De Witt, B. S. and Graham, R. D. (eds.) (1973). *The many-worlds interpretation of quantum mechanics*. Princeton University Press.

Дельбрюк [1986] Delbrück, M. (1986). *Mind from matter?* Blackwell Scientific Publishing, Oxford.

Деннетт [1978] Dennett, D. C. (1978). *Brainstorms*. Philosophical Essays on Mind and Psychology, Harvester Press, Hassocks, Sussex.

Джейнс [1980] Jaynes, J. (1980). *The origin of consciousness in the breakdown of the bicameral mind*. Penguin Books Ltd., Harmondsworth, Middx.

Дике, Грётцингер, Корнхубер [1976] Deeke, L., Grotzinger, B., and Kornhuber, H. H. (1976). Voluntary finger movements in man: cerebral potentials and theory. *Biol. Cybernetics*, **23**, 99.

Дирак [1928] Dirac, P. A. M. (1928). The quantum theory of the electron. *Proc. Roy. Soc.* (*Lond.*), **A117**, 610–24; *ditto*, part II, *ibid.*, **A118**, 361. (Рус. пер.: *Дирак П. А. М.* Квантовая теория электрона // Дирак П. А. М. К созданию квантовой теории поля. М.: Наука, 1990. Ч. 1, 2. С. 113–141.)

Дирак [1938] Dirac, P. A. M. (1938). Classical theory of radiating electrons. *Proc. Roy. Soc.* (*Lond.*), **A167**, 148.

Дирак [1939] Dirac, P. A. M. (1939). The relations between mathematics and physics. *Proc. Roy. Soc., Edinburgh*, **59**, 122. (Рус. пер.: *Дирак П. А. М.* Отношение между математикой и физикой // Дирак П. А. М. К созданию квантовой теории поля. С. 245–254.)

Дирак [1947] Dirac, P. A. M. (1947). *The principles of quantum mechanics* (3rd edn). Oxford University Press. (Рус. пер.: *Дирак П. А. М.* Принципы квантовой механики. М.: Наука, 1979.)

Дирак [1982] Dirac, P. A. M. (1982). Pretty mathematics. *Int. J. Theor. Phys.*, **21**, 603–5.

Дойч [1985] Deutsch, D. (1985). Quantum theory, the Church-Turing principle and the universal quantum computer. *Proc. Roy. Soc.* (*Lond.*), **A400**, 97–117.

Доукинс [1986] Dawkins, R. (1986). *The blind watchmaker.* Longman, London.

Дрэйк [1953] Drake, S. (trans.) (1953). *Galileo Galilei: dialogue concerning the two chief world systems — Ptolemaic and Copernican.* University of California, Berkeley, 1953.

Дрэйк [1957] Drake, S. (1957). *Discoveries and opinions of Galileo.* Doubleday, New York.

Дэвис [1974] Davies, P. C. W. (1974). *The physics of time-asymmetry.* Surrey University Press.

Дэвис [1988] Davis, M. (1988). Mathematical logic and the origin of modern computers. Опубликовано в *The universal Turing machine: a half-century survey* (ed. R. Herken), Kammerer & Unverzagt, Hamburg.

Дэвис и др. [1987] Davies, R. D., Lasenby, A. N., Watson, R. A., Daintree, E. J., Hopkins, J., Beckman, J., Sanchez-Almeida, J., and Rebolo, R. (1987). Sensitive measurement of fluctuations in the cosmic microwave background. *Nature*, **326**, 462–5.

Дэвис, Браун [1988] Davies, P. C. W. and Brown, J. (1988). *Superstrings: a theory of everything?* Cambridge University Press.

Дэвлин [1988] Devlin, K. (1988). *Mathematics: the new golden age.* Penguin Books, London.

Кандел [1976] Kandel, E. R. (1976). *The cellular basis of behaviour.* Freeman, San Francisco.

Каройхази [1974] Károlyházy, F. (1974). Gravitation and quantum mechanics of macroscopic bodies. *Magyar Fizikai Folyóirat*, **12**, 24.

Каройхази, Френкель, Лукач [1986] Károlyházy, F., Frenkel, A., and Lukács, B. (1986). On the possible role of gravity on the reduction of the wave function. Опубликовано в *Quantum concepts in space and time* (ed. R. Penrose and C. J. Isham), Oxford University Press.

Картан [1923] Cartan, É. (1923). Sur les variétés à connexion affine et la théorie de la relativité généralisée. *Ann. Sci. Ec. Norm. Sup.*, **40**, 325–412.

Катленд [1980] Cutland, N. J. (1980). *Computability: an introduction to recursive function theory.* Cambridge University Press.

Кин [1988] Keene, R. (1988). Chess: Henceforward. *The Spectator*, **261** (№ 8371), 52.

Клаузер, Хорн, Шимони, Холт [1969] Clauser, J. F., Horne, A. H., Shimony, A., and Holt, R. A. (1969). Proposed experiment to test local hidden-variable theories. Опубликовано в *Quantum theory and measurement* (ed. J. A. Wheeler and W. H. Zurek), Princeton University Press, 1983; первоначально опубликовано в *Phys. Rev. Lett.*, **23**, 880–4.

Клос [1983] Close, F. (1983). *The cosmic onion: quarks and the nature of the universe.* Heinemann, London.

Кнут [1981] Knuth, D. M. (1981). *The art of computer programming*, Vol. 2 (2nd edn). Addison-Wesley, Reading, MA. (Рус. пер.: *Кнут Д.* Искусство программирования. М.: Издательский дом «Вильямс», 2000.)

Комар [1964] Komar, A. B. (1964). Undecidability of macroscopically distinguishable states in quantum field theory. *Phys. Rev.*, **133B**, 542–4.

Комар [1969] Komar, A. B. (1969). Qualitative features of quantized gravitation. *Int. J. Theor. Phys.* **2**, 157–60.

Коэн [1966] Cohen, P. C. (1966). *Set theory and the continuum hypothesis.* Benjamin, Menlo Park, CA.

Кузнецов [1980] *Кузнецов Б. Г.* Эйнштейн: Жизнь. Смерть. Бессмертие. М.: Наука, 1980.

Леви [1984] Levy, D. W. L. (1984). *Chess computer handbook.* Batsford.

Леду [1985] LeDoux, J. E. (1985). Brain, mind and language. Опубликовано в *Brain and mind* (ed. D. A. Oakley), Methuen, London and New York.

Либет [1987] Libet, B. (1987). Consciousness: Conscious subjective experience. Опубликовано в *Encyclopedia of neuroscience*, Vol. 1 (ed.) G. Adelman. Birkhauser; pp. 271–5.

Либет [1989] Libet, B. (1989). Conscious subjective experience vs. unconscious mental functions: A theory of the cerebral process involved. Опубликовано в *Models of brain function* (ed. R. M. J. Cotterill), Cambridge University Press, Cambridge; pp. 35–43.

Либет, Райт, Файнштейн, Перл [1979] Libet, B., Wright, E. W. Jr., Feinstein, B., and Pearl, D. K. (1979). Subjective referral of the timing for a conscious sensory experience. *Brain*, **102**, 193–224.

Лоренц [1972] Lorenz, K. (1972). Цитируется по *From ape to Adam*, by H. Wendt, Bobbs Merrill, Indianapolis.

Лукас [1961] Lucas, J. R. (1961). Minds, machines and Gödel. *Philosophy*, **36**, 120–4. Перепечатано в Alan Ross Anderson (1964), *Minds and machines*, Englewood Cliffs.

Льюис [1969] Lewis, D. (1969). Lucas against mechanism. *Philosophy*, **44**, 231–3.

Льюис [1989] Lewis, D. (1989). Lucas against mechanism II. *Can. J. Philos.* **9**, 373–6.

Майерс [1974] Myers, D. (1974). Nonrecursive tilings of the plane, II. *J. Symbolic Logic*, **39**, 286–94.

Майорана [1932] Majorana, E. (1932). Atomi orientati in campo magnetico variabile. *Nuovo Cimento*, **9**, 43–50.

МакКей [1987] MacKay, D. (1987). Divided brains — divided minds? Опубликовано в *Mindwaves* (ed. C. Blakemore and S. Greenfield), Basil Blackwell, Oxford.

Максвелл [1865] Maxwell, J. C. (1865). A dynamical theory of the electromagnetic field. *Philos. Trans. Roy. Soc. (Lond.)*, **155**, 459–512. (Рус. пер.: *Максвелл Дж. К.* Динамическая теория поля. Ч. VI. Электромагнитная теория света. Избр. соч. по теории электромагнитного поля: М.: Гостехтеориздат, 1952.)

Мандельброт [1986] Mandelbrot, B. B. (1986). Fractals and the rebirth of iteration theory. Опубликовано в *The beauty of fractals: images of complex dynamical systems*, H.-O. Peitgen and P. H. Richter, Springer-Verlag, Berlin; pp. 151–60. (Рус. пер.: *Мандельброт Б.* Фракталы и возрождение теории иттераций // Пайтген X. О., Рихтер П. X. Красота фракталов. М.: Мир, 1993.)

Мандельброт [1989] Mandelbrot, B. B. (1989). Some 'facts' that evaporate upon examination. *Math. Intelligencer*, **11**, 12–16.

Мермин [1985] Mermin, D. (1985). Is the moon there when nobody looks? Reality and the quantum theory. *Physics Today*, **38** (№ 4), 38–47.

Мизнер [1969] Misner, C. W. (1969). Mixmaster universe. *Phys. Rev. Lett.*, **22**, 1071–4.

Мики [1988] Michie, D. (1988). The fifth generation's unbridged gap. Опубликовано в *The universal Turing machine: a half-century survey* (ed. R. Herken), Kammerer & Unverzagt, Hamburg.

Мински [1968] Minsky, M. L. (1968). Matter, mind, and models. Опубликовано в *Semantic information processing.* (ed. M. L. Minsky), MIT Press, Cambridge, Mass.

Мирс, Сперри [1953] Myers, R. E. and Sperry, R. W. (1953). Interocular transfer of a visual form discrimination habit in cats after section of the optic chiasm and corpus callosum. *Anatomical Record*, **175**, 351–2.

Моравец [1989] Moravec, H. (1989). *Mind children: the future of robot and human intelligence.* Harvard University Press.

Моруцци, Магун [1949] Moruzzi, G. and Magoun, H. W. (1949). Brainstem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology*, **1**, 455–73.

Мотт [1929] Mott, N. F. (1929). The wave mechanics of α-ray tracks. Опубликовано в *Quantum theory and measurement* (ed. J. A. Wheeler and W. H. Zurek), Princeton University Press, 1983; первоначально опубликовано в *Proc. Roy. Soc.* (*Lond.*), **A126**, 79–84.

Мотт, Месси [1965] Mott, N. F. and Massey, H.S.W. (1965). *The theory of atomic collisions.* Clarendon Press, Oxford. (Рус. пер.: *Мотт Н., Месси Г.* Теория атомных столкновений. М.: Мир, 1969.)

Нагель, Ньюман [1958] Nagel, E. and Newman, J. R. (1958). *Gödel's proof.* Routledge & Kegan Paul Ltd.

Нельсон, Халперин [1985] Nelson, D. R. and Halperin, B. I. (1985). Pentagonal and icosahedral order in rapidly cooled metals. *Science*, **229**, 233.

Ньютон [1687] Newton, I. (1687). *Principia.* Cambridge University Press. (Рус. пер.: *Ньютон И.* Математические начала натуральной философии / Пер. с лат. А. И. Крылова. Петроград, 1916 (соврем. изд.: М.: Издательство ЛКИ/URSS, 2008.))

Ньютон [1730] Newton, I. (1730). *Opticks.* 1952, Dover, Inc. (Рус. пер.: *Ньютон И.* Оптика. М.—Л., 1954.)

О'Кифи [1985] O'Keefe, J. (1985). Is consciousness the gateway to the hippocampal cognitive map? A speculative essay on the neural basis of mind. Опубликовано в *Brain and mind* (ed. D. A. Oakley), Methuen, London and New York.

О'Коннелл [1988] O'Connell, K. (1988). Computer chess. *Chess*, **15**.

Окли [1985] Oakley, D. A. (ed.) (1985). *Brain and mind.* Methuen, London and New York.

Окли, Имз [1985] Oakley, D. A. and Eames, L. C. (1985). The plurality of consciousness. Опубликовано в *Brain and mind* (ed. D. A. Oakley), Methuen, London and New York.

Онода, Стайнхардт, Ди Винченцо, Соколар [1988] Onoda, G. Y., Steinhardt, P. J., DiVincenzo, D. P., and Socolar, J. E. S. (1988). Growing perfect quasicrystals. *Phys. Rev. Lett.*, 60, 2688.

Оппенгеймер, Снайдер [1939] Oppenheimer, J. R. and Snyder, H. (1939). On continued gravitational contraction. *Phys. Rev.* **56**, 455–9.

Пайс [1982] Pais, A. (1982). *'Subtle is the Lord...': the science and the life of Albert Einstein.* Clarendon Press, Oxford.

Пайтген, Заупе [1988] Peitgen, H.-O. and Saupe, D. (1988). *The science of fractal images.* Springer-Veriag, Berlin.

Пайтген, Рихтер [1986] Peitgen, H.-O. and Richter, P. H. (1986). *The beauty of fractals.* Springer-Verlag, Berlin and Heidelberg. (Рус. пер.: *Пайтген Х.-О., Рихтер П. Х.* Красота фракталов. М.: Мир, 1993.)

Парис, Харрингтон [1977] Paris, J. and Harrington, L. (1977). A mathematical incompleteness in Peano arithmetic. Опубликовано в *Handbook of mathematical logic* (ed. J. Barwise), North-Holland, Amsterdam.

Пенроуз [1965] Penrose, R. (1965). Gravitational collapse and space-time singularities. *Phys. Rev. Lett.*, **14**, 57–9. (Рус. пер.: *Пенроуз Р.* Гравитационный коллапс и пространственно-временные сингулярности // Альберт Эйнштейн и теория гравитации. М.: Мир, 1979. С. 390–395.)

Пенроуз [1974] Penrose, R. (1974). The role of aesthetics in pure and applied mathematical research. *Bull. Inst. Math. Applications*, **10**, № 7/8, 266–71.

Пенроуз [1979a] Penrose, R. (1979a). Einstein's vision and the mathematics of the natural world. *The Sciences* (*March*), 6–9.

Пенроуз [1979б] Penrose, R. (1979b). Singularities and time-asymmetry. Опубликовано в *General relativity: An Einstein centenary* (ed. S. W. Hawking and W. Israel), Cambridge University Press. (Рус. пер.: *Пенроуз Р.* Сингулярности и асимметрия во времени // Общая теория относительности. М.: Мир, 1983. С. 233–295.)

Пенроуз [1987a] Penrose, R. (1987a). Newton, quantum theory and reality. Опубликовано в *300 years of gravity* (ed. S. W. Hawking and W. Israel), Cambridge University Press.

Пенроуз [1987б] Penrose, R. (1987b). Quantum Physics and Conscious Thought. Опубликовано в *Quantum implications: Essays in honour of David Bohm* (ed. B. J. Hiley and F. D. Peat), Routledge and Kegan Paul, London & New York.

Пенроуз [1989a] Penrose, R. (1989a). Tilings and quasi-crystals; a non-local growth problem? Опубликовано в *Aperiodicity and order* 2 (ed. M. Jarič), Academic Press, New York.

Пенроуз [1989б] Penrose, R. (1989b). Difficulties with inflationary cosmology. Опубликовано в *The Fourteenth Texas Symposium on Relativistic Astrophysics* (ed. E. J. Fenyves), N. Y. Acad. Sci., New York, **571**, 249–64.

Пенроуз, Риндлер [1984] Penrose, R. and Rindler, W. (1984). *Spinors and space-time*, Vol. 1: *Two-spinor calculus and relativistic fields.* Cambridge University Press. (Рус. пер.: *Пенроуз Р., Риндлер В.* Спиноры и пространство-время. М.: Мир, 1988.)

Пенроуз, Риндлер [1986] Penrose, R. and Rindler, W. (1986). *Spinors and space-time*, Vol. 2: *Spinor and twistor methods in space-time geometry.* Cambridge University Press. (Рус. пер.: Там же.)

Пенфилд, Джаспер [1947] Penfield, W. and Jasper H. (1947). Highest level seizures. *Research Publications*

of the Association for Research in Nervous and Mental Diseases (*New York*), **26**, 252–71.

Перл [1985] Pearle, P. (1985). 'Models for reduction'. Опубликовано в *Quantum concepts in space and time* (ed. C. J. Isham and R. Penrose), Oxford University Press.

Перл [1989] Pearle, P. (1989). Combining stochastic dynamical state-vector reduction with spontaneous localization. *Phys. Rev. A*, **39**, 2277–89.

Пур-Эль, Ричардс [1979] Pour-El, M. B. and Richards, I. (1979). A computable ordinary differential equation which possesses no computable solution. *Ann. Math. Logic*, **17**, 61–90.

Пур-Эль, Ричардс [1981] Pour-El, M. B. and Richards, I. (1981). The wave equation with computable initial data such that its unique solution is not computable. *Adv. in Math.*, **39**, 215–39.

Пур-Эль, Ричардс [1982] Pour-El, M. B. and Richards, I. (1982). Noncomputability in models of physical phenomena. *Int. J. Theor. Phys.*, **21**, 553–5.

Пур-Эль, Ричардс [1989] Pour-El, M. B. and Richards, I. (1989). *Computability in analysis and physics*. Springer-Verlag, New York.

Ракер [1984] Rucker, R. (1984). *Infinity and the mind: the science and philosophy of the infinite*. Paladin Books, Granada Publishing Ltd., London (first published by Birkhauser Inc., Boston, Mass., 1982.).

Рауз Болл [1892] Rouse Ball, W. W. (1892). Calculating prodigies. Опубликовано в *Mathematical recreations and essays*.

Резников, Уэллс [1984] Resnikoff, H. L. and Wells, R. O. Jr. (1973). *Mathematics and civilization*. Holt, Rinehart and Winston, Inc., New York; перепечатано с дополнениями (1984) Dover Publications, Inc., Mineola, New York.

Риндлер [1977] Rindler, W. (1977). *Essential relativity*. Springer-Verlag, New York.

Риндлер [1982] Rindler, W. (1982). *Introduction to special relativity*. Clarendon Press, Oxford.

Робинсон [1971] Robinson, R. M. (1971). Undecidability and nonperiodicity for tilings of the plane. *Invent. Math.*, **12**, 177–209.

Рэй [1986] Rae, A. (1986). *Quantum physics: illusion or reality?* Cambridge University Press.

Сакс [1962] Sachs, R. K. (1962). Gravitational waves in general relativity. VIII. Waves in asymptotically flat space-time. *Proc. Roy. Soc. London*, **A270**, 103–26.

Серл [1980] Searle, J. (1980). Minds, brains and programs. Опубликовано в *The behavioral and brain sciences*, Vol. 3. Cambridge University Press. Перепечатано в *The mind's I* (ed. D. R. Hofstadter and D. C. Dennett), Basic Books, Inc., Penguin Books Ltd., Harmondsworth, Middx., 1981.

Серл [1987] Searle, J. R. (1987). Minds and brains without programs. Опубликовано в *Mindwaves* (ed. C. Blakemore and S. Greenfield). Basil Blackwell, Oxford.

Сквайерс [1985] Squires, E. (1985). *To acknowledge the wonder*. Adam Hilger Ltd., Bristol.

Сквайерс [1986] Squires, E. (1986). *The mystery of the quantum world*. Adam Hilger Ltd., Bristol.

Смит [1983] Smith, S. B. (1983). *The great mental calculators*. Columbia University Press.

Сморински [1983] Smorynski, C. (1983). 'Big' news from Archimedes to Friedman. *Notices Amer. Math. Soc.*, 30, 251–6.

Сперри [1966] Sperry, R. W. (1966). Brain bisection and consciousness. Опубликовано в *Brain and conscious experience* (ed. J. C. Eccles), Springer, New York.

Типлер, Кларк, Эллис [1980] Tipler, F. J., Clarke, C. J. S., and Ellis, G. F. R. (1980). Singularities and horizons — a review article. Опубликовано в *General relativity and gravitation* (ed. A. Held), Vol. 2, pp. 97–206. Plenum Press, New York.

Трейман, Джекив, Зумино, Виттен [1985] Treiman, S. B., Jackiw, R, Zumino, B., and Witten, E. (1985). *Current algebra and anomalies, Princeton series in physics*. Princeton University Press, Princeton, N. J.

Тьюринг [1937] Turing, A. M. (1937). On computable numbers, with an application to the Entscheidungsproblem. *Proc. Land. Math. Soc.* (*ser.* 2), **42**, 230–65; a correction **43**, 544–6.

Тьюринг [1939] Turing, A. M. (1939). Systems of logic based on ordinals. *P. Lond. Math. Soc.*, **45**, 161–228.

Тьюринг [1950] Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, **59**, № 236. Перепечатано в *The mind's I* (ed. D. R. Hofstadter and D. C. Dennett), Basic Books, Inc.; Penguin Books, Ltd., Harmondsworth, Middx., 1981.

Уилер [1983] Wheeler, J. A. (1983). Law without law. Опубликовано в *Quantum theory and measurement* (ed. J. A. Wheeler and W. H. Zurek), Princeton University Press, pp. 182–213.

Уилер, Фейнман [1945] Wheeler, J. A. and Feynman, R. P. (1945). Interaction with the absorber as the mechanism of radiation. *Revs. Mod. Phys.*, **17**, 157–81.

Уилер, Цурек [1983] Wheeler, J. A. and Zurek, W. H. (eds.) (1983). *Quantum theory and measurement*. Princeton University Press.

Уилл [1987] Will, C. M. (1987). Experimental gravitation from Newton's *Principia* to Einstein's general relativity. Опубликовано в 300 *years of gravitation* (ed. S. W. Hawking and W. Israel), Cambridge University Press.

Уиттекер [1910] Whittaker, E. T. (1910). *The history of the theories of aether and electricity*. Longman, London. (Рус. пер.: *Уиттекер Э.* История теории эфира и электричества. Ижевск: РХД, 2001.)

Уолтер [1953] Grey Walter, W. (1953). *The living brain*. Gerald Duckworth and Co. Ltd.

Уорд, Уэллс [1990] Ward, R. S. and Wells R. O. Jr. (1990). Twistor geometry and field theory. Cambridge University Press.

Фейнман [1985] Feynman, R. P. (1985). *QED: the strange theory of light and matter*. Princeton University Press. (Рус. пер.: *Фейнман Р.* КЭД, странная теория света и вещества. М.: Наука, 1988.)

Фейнман, Лейтон, Сэндс [1965] Feynman, R. P., Leighton, R. B., and Sands, M. (1965). *The Feynman Lectures*. Addison-Wesley. (Рус. пер.: *Фейнман Р., Лейтон Р., Сэндс М.* Фейнмановские лекции по физике. М.: Мир, 1977.)

Феферман [1988] Feferman, S. (1988). Turing in the Land of O(z). Опубликовано в *The universal Turing machine: a half-century survey* (ed. R. Herken), Kammerer & Unverzagt, Hamburg.

Фодор [1983] Fodor, J. A. (1983). *The modularity of mind.* MIT Press, Cambridge, Mass.

фон Нейман [1955] von Neumann, J. (1955). *Mathematical foundations of quantum mechanics.* Princeton University Press. (Рус. пер.: *Нейман И.* Математические основы квантовой механики. М.: Наука, 1964.)

Фредкин, Тоффоли [1982] Fredkin, E. and Toffoli, T. (1982). Conservative logic. *Int. J. Theor. Phys.*, **21**, 219–53.

Фридман, Клаузер [1972] Freedman, S. J. and Clauser, J. F. (1972). Experimental test of local hidden-variable theories.Опубликовано в *Quantum theory and measurement* (ed. J. A. Wheeler and W. H. Zurek), Princeton University Press, 1983; первоначально опубликовано в *Phys. Rev. Lett.*, **28**, 938–41.

Хаггетт, Тод [1985] Huggett, S. A. and Tod, K. P. (1985). *An introduction to twistor theory.* London Math. Soc. student texts, Cambridge University Press.

Хайли, Пит [1987] Hiley, B. J. and Peat, F. D. (eds.) (1987). *Quantum implications. Essays in honour of David Bohm.* Routledge and Kegan Paul, London & New York.

Ханф [1974] Hanf, W. (1974). Nonrecursive tilings of the plane, I. *J. Symbolic Logic*, **39**, 283–5.

Харт [1982] Harth, E. (1982). *Windows on the mind.* Harvester Press, Hassocks, Sussex.

Хартли, Хокинг [1983] Hartle, J. B. and Hawking, S. W. (1983). Wave function of the universe. *Phys. Rev.*, **D31**, 1777.

Хебб [1954] Hebb, D. O. (1954). The problem of consciousness and introspection. Опубликовано в *Brain mechanisms and consciousness* (ed. J. F. Delafresnaye), Blackwell, Oxford.

Херкен [1988] Herken, R. (ed.) (1988). *The universal Turing machine: a half-century survey.* Kammerer & Unverzagt, Hamburg.

Хехт, Шлейер, Пирэн [1941] Hecht, S., Shlaer, S and Pirenne, M. H. (1941). Energy, quanta and vision. *J. of Gen. Physiol*, **25**, 891–40.

Ходжис [1983] Hodges, A. P. (1983). *Alan Turing: the enigma.* Burnett Books and Hutchinson, London; Simon and Schuster, New York.

Хокинг [1975] Hawking, S. W. (1975). Particle creation by black holes. *Commun. Math. Phys.*, **43**, 199–220. (Рус. пер.: *Хокинг С.* Рождение частиц на черных дырах // Альберт Эйнштейн и теория гравитации. М.: Мир, 1979. С. 479–510.)

Хокинг [1987] Hawking, S. W. (1987). Quantum cosmology. Опубликовано в 300 *years of gravitation* (ed. S. W. Hawking and W. Israel), Cambridge University Press.

Хокинг [1988] Hawking, S. W. (1988). *A brief history of time.* Bantam Press, London. (Рус. пер.: *Хокинг С.* Краткая история времени. СПб.: Амфора, 2000.)

Хокинг, Пенроуз [1970] Hawking, S. W. and Penrose, R. (1970). The singularities of gravitational collapse and cosmology. *Proc. Roy. Soc.* (*London*), **A314**, 529–48.

Хофштадтер [1979] Hofstadter, D. R. (1979). *Gödel, Escher, Bach: an eternal golden braid.* Harvester Press, Hassocks, Sussex. (Рус. пер.: *Хофштадтер Д.* Гедель, Эшер, Бах: эта бесконечная гирлянда. Самара: Издательский дом «Бахрах-М», 2001.)

Хофштадтер [1981] Hofstadter, D. R. (1981). A conversation with Einstein's brain. Опубликовано в *The mind's I* (ed. D. R. Hofstadter and D. C. Dennett), Basic Books, Inc.; Penguin Books, Ltd., Harmondsworth, Middx.

Хофштадтер, Деннетт [1981] Hofstadter, D. R. and Dennett, D. C. (eds.) (1981). *The mind's I.* Basic Books, Inc.; Penguin Books, Ltd., Harmondsworth, Middx.

Хьюбел [1988] Hubel, D. H. (1988). *Eye, brain and vision.* Scientific American Library Series #22.

Чандрасекар [1987] Chandrasekhar, S. (1987). *Truth and beauty: aesthetics and motivations in science.* University of Chicago Press.

Черч [1941] Church, A. (1941). *The calculi of lambda-conversion.* Annals of Mathematics Studies, № 6. Princeton University Press.

Черчланд [1984] Churchland, P. M. (1984). *Matter and consciousness.* Bradford Books, MIT Press, Cambridge, Mass.

Шенк, Абельсон [1977] Schank, R. C. and Abelson, R. P. (1977). *Scripts, plans, goals and understanding.* Erlbaum, Hillsdale, N. J.

Шехтман, Блех, Гратиас, Кан [1984] Shechtman, D., Blech, I., Gratias, D., and Cahn, J. W. (1984). Metallic phase with long-range orientational order and no translational symmetry. *Phys. Rev. Lett.*, **53**, 1951.

Шредингер [1935] Schrödinger, E. (1935). Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften*, **23**, 807–12, 823–8, 844–9. (Translation by J. T. Trimmer (1980). Опубликовано в *Proc. Amer. Phil. Soc.*, **124**, 323–38.) Опубликовано в *Quantum theory and measurement* (ed. J. A. Wheeler and W. H. Zurek), Princeton University Press, 1983.

Шредингер [1967] Schrödinger, E. (1967). *'What is life?' and 'Mind and matter'.* Cambridge University Press. (Рус. пер.: *Шредингер Э.* Что такое жизнь? С точки зрения физика. 2-е изд. М.: Атомиздат, 1972.)

Эверетт [1957] Everett, H. (1957). 'Relative state' formulation of quantum mechanics. Опубликовано в *Quantum theory and measurement* (ed. J. A. Wheeler and W. H. Zurek), Princeton University Press, 1983; первоначально опубликовано в *Rev. of Mod. Phys.*, **29**, 454–62.

Эйнштейн, Подольский, Розен [1935] Einstein, A., Podolsky, B., and Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? Опубликовано в *Quantum theory and measurement* (ed. J. A. Wheeler and W. H. Zurek), Princeton University Press, 1983; первоначально опубликовано в *Phys. Rev.*, **47**, 777–80.

Экклз [1973] Eccles, J. C. (1973). *The understanding of the brain.* McGraw-Hill, New York.

Эткинс [1987] Atkins, P. W. (1987). *Why mathematics works.* Oxford University Extension Lecture in series: Philosophy and the New Physics (13 March).

Иллюстративный материал, используемый в книге

- Рис. 4.6 и 4.9. Воспроизводится из *The Mathematical Gardner* (D. A. Klarner (ed.)) Wadsworth International, 1981.
- Рис. 4.7. Воспроизводится из *Tilings and Patterns* (B. Grünbaum and G. C. Shephard) © W. H. Freeman, 1987.
- Рис. 4.10. Воспроизводится из *Hermann Weyl* 1885–1985 (K. Chandrasekharan) Springer, 1986.
- Рис. 4.11 и 10.3. Воспроизводится из Pentalplexity: a class of non-periodic tilings of the plane. *The Mathematical Intelligence*, 2, 32–7 (Springer, 1979).
- Рис. 4.12. Воспроизводится из *M. C. Escher: Art and Science* (H. S. M. Coxeter, M. Emmer, R. Penrose and M. L. Teuber (eds)) North-Holland, 1986.
- Рис. 5.2. Воспроизводится из *M. C. Escher: Heirs* © 2002 *Cordon Art B. V. — Baarn — Holland*. Все права защищены.
- Рис. 10.4. Воспроизводится из *Journal of Materials Research*, 2, 1–4 (Materials Research Society, 1987).

Все остальные рисунки (включая 4.10 и 4.12) принадлежат автору.

Именной указатель

Предметный указатель

Оглавление

Серия «Синергетика: от прошлого к будущему»

Нельзя построить содержательную общую теорию нелинейных систем, – считал Джон фон Нейман.

Великий математик ошибался.

В этом убеждают книги этой серии, посвященные синергетической парадигме, нелинейной науке, бифуркациям, фракталам, хаосу и многим другим интересным вещам.



Выдающийся ученый современности, активно работающий в различных областях математики, общей теории относительности и квантовой теории; автор теории твисторов.

Р. Пенроуз возглавляет кафедру математики Оксфордского университета, а также является почетным профессором многих зарубежных университетов и академий. Он является членом Лондонского королевского общества. Среди его наград — премия Вольфа (совместно с С. Хокингом), медаль Дирака, премия Альберта Эйнштейна и медаль Королевского общества. В 1994 г. за выдающиеся заслуги в развитии науки королевой Англии ему был присвоен титул *сэра*.

Roger PENROSE

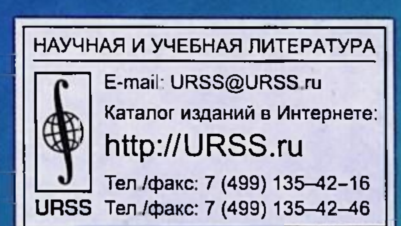
THE EMPEROR'S NEW MIND

Concerning Computers, Minds and The Laws of Physics

- Может ли компьютер обладать разумом?
- Алгоритмы и машины Тьюринга
- Математика и действительность
- Истина, доказательство и интуиция
- Классический мир
- Квантовая магия и квантовое таинство
- Космология и стрела времени
- В поисках квантовой теории гравитации
- Реальный мозг и модели мозга
- Где находится физика ума?


5955 ID 75468

9 785382 007441 >


НАУЧНАЯ И УЧЕБНАЯ ЛИТЕРАТУРА
URSS
E-mail: URSS@URSS.ru
Каталог изданий в Интернете:
http://URSS.ru
Тел./факс: 7 (499) 135–42–16
Тел./факс: 7 (499) 135–42–46